Title
A Logical Theory of Confirmation

Author
Ahmed-Buehler, Russell Joseph

Publication Date
2019

Peer reviewed|Thesis/dissertation

A Logical Theory of Confirmation

By

Russell Joseph Ahmed-Buehler

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Logic and the Methodology of Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Associate Professor Lara Buchak, Chair
Associate Professor Wesley Holliday
Professor John MacFarlane
Associate Professor Seth Yalcin
Associate Professor Shamik Dasgupta

Summer 2019

Abstract

A Logical Theory of Confirmation

by

Russell Joseph Ahmed-Buehler

Doctor of Philosophy in Logic and the Methodology of Science

University of California, Berkeley

Associate Professor Lara Buchak, Chair

This dissertation concerns the interpretation and structure of two intuitive notions: rational credence and confirmation. Probabilistic accounts of rational credence currently enjoy a position of considerable prestige, underwriting significant work not only in philosophy but also in economics and statistics. Confirmation, in contrast, is widely regarded as ill-formed, a misleading misconception akin to phlogiston, witches, or cosmic ether (de Finetti 1979, x). The primary project of this dissertation is to undercut the contemporary consensus on both notions by first demonstrating a systemic weakness in probabilistic accounts of rational credence (part I) and then providing a non-probabilistic account of confirmation (part II). Since any adequate account of confirmation is *prima facie* an adequate account of rational credence, the negative work of part one dovetails with the positive account offered in part two. Confirmation is no misleading misconception, and rational credences are not probability functions.

Contents

Acknowledgments

This dissertation would not have been possible without the large number of mentors, friends, and family members who have supported my work. I would first like to thank the members of my committee: Lara Buchak who waded through countless drafts and always had time to help, Wes Holliday who guided so much of my early work and has always provided insightful comments, Seth Yalcin and Shamik Dasgupta who pushed me to expand my views in unexpected ways, and John MacFarlane whose gracious support over the last few months made it all possible. I would also like to thank John Campbell, Tim Clarke, and Paolo Mancosu in particular from amongst the faculty at Berkeley as well as Jason Decker from the faculty at Carleton. All of you have had a profound influence on both my life and my work, and I am deeply thankful to have been your student.

Second, I would like to thank the many graduate students in the department for their support and friendship. While the full list is rather long, I'd like to particularly thank Justin Bledin, Ravit Dotan, Melissa Fusco, Matthew Harrison-Trainor, Jeff Kaplan, Richard Lawrence, Rachel Rudolph, Justin Vlasits, James Walsh, and Michael Wan. Special thanks to Sven Neth and Mikayla Kelley for suffering through early drafts of my work and providing helpful feedback. Extra special thanks to Arc Kocurek without whom my time at Berkeley would have been much diminished.

Third and finally, I would like to thank my family. Thank you to my wonderful wife Sameena for her endless encouragement as well as to my parents, Kim and Fred, for their unwavering support over the last 29 years. Thank you to my in-laws—Sheriff, Bridget, Sharmila, Peter, Sharia, and Josh—for your advice, for your support, and for occasionally pretending to care about logic. Thank you to my brother Matt and sister-in-law Hailee for making sure I don't forget how to relax and have fun. I could not have done it without all of you.

Chapter 1

Introduction

This dissertation concerns the interpretation and structure of two intuitive notions: rational credence and confirmation. Probabilistic accounts of rational credence currently enjoy a position of considerable prestige, underwriting significant work not only in philosophy but also in economics and statistics. Confirmation, in contrast, is widely regarded as ill-formed, a misleading misconception akin to phlogiston, witches, or cosmic ether (de Finetti 1979, x). The primary project of this dissertation is to undercut the contemporary consensus on both notions by first demonstrating a systemic weakness in probabilistic accounts of rational credence (part I) and then providing a non-probabilistic account of confirmation (part II). The current chapter provides a brief introduction to probability, rational credence, and confirmation as well as an outline of the rest of the dissertation.

1.1 Probability

In contemporary mathematics, a probability space is a triple $\langle \Omega, \mathcal{F}, \mu \rangle$ where Ω is a non-empty set ("the state space"), $\mathcal{F}$ is an appropriate[1] collection of subsets of Ω ("the event space"), and μ is a function from $\mathcal{F}$ to the real numbers $\mathbb{R}$ ("the probability function") satisfying the following three criteria:

(K1) $\mu(\Omega) = 1$;

(K2) For any $A \in \mathcal{F}$, $\mu(A) \geq 0$;

(K3) For a countable collection $\{A_i\}$ of pairwise disjoint sets from $\mathcal{F}$,

$$\mu(\cup\{A_i\}) = \sum_i \mu(A_i).$$

[1] More exactly, $\mathcal{F}$ is a σ-algebra, a subset of $\mathcal{P}(\Omega)$ containing $\emptyset$ that is closed under complement, countable union, and countable intersection.

Informally, μ assigns the trivially true event probability one, the trivially false event probability zero, and all other events values between these two (inclusive). In addition, μ combines the weights of incompatible (i.e., disjoint) events by adding their respective values.

A die, for example, naturally gives rise to a state space Ω containing a state for each numbered face of the die.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

The event space $\mathcal{F}$ is then typically taken to be the set of all subsets of this state space. This includes not only singleton events—e.g., the die shows six—but also any combination of these singletons, e.g., the die shows six or the die shows four.

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \dots, \{1, 2, 3, 4, 5, 6\}\}$$

Finally, dice are conventionally associated with the probability function μ which maps each singleton event to the value $\frac{1}{6}$ and combination events to the sum of the corresponding singleton events.

$$\begin{aligned}
\mu(\emptyset) &= 0 \\
\mu(\{1\}) &= \frac{1}{6} \\
\mu(\{2\}) &= \frac{1}{6} \\
\mu(\{3\}) &= \frac{1}{6} \\
&\vdots \\
\mu(\{1, 2, 3, 4, 5, 6\}) &= 1
\end{aligned}$$

Putting each of these components together, the triple $\langle \Omega, \mathcal{F}, \mu \rangle$ is a probability space.

The contemporary meaning of ‘probability’ occupies an uncomfortable middle ground between this mathematical definition and an unanalyzed notion of likelihood. This ambiguity is intolerable in the current setting, and I have found it easiest to explicitly adopt the former, mathematical usage. There is thus no sense in asking why probability functions should be additive or why the real numbers are the right choice for probability. There is also, however, no special connection between the probability space described above, the uncertainty of rolling the die in question, and the likelihood of a particular outcome. μ is just as legitimate as an assignment of areas to each face of the die—$\frac{1}{6}$ square inches each!—as it is a measure of likelihood. If probability should happen to be an adequate formal analysis of likelihood, so much the better. If it should not, then nothing has been lost. Probability is all and only what satisfies the mathematical definition provided.

While this thoroughly formal use of ‘probability’ breaks with some commentaries, it is in line with the general shift towards axiomatization in mathematics and its usage there. The fact that mathematical work on probability originates with games of chance and notions of likelihood is interesting but ultimately immaterial. As Hilbert himself is supposed to have remarked about geometry, "One must be able to say at all times—instead of points, straight lines, and planes—tables, chairs, and beer mugs" (Reid and Weyl 2013, 57). Probability functions and spaces are understood here in the same spirit.

1.2 Rational Credence

We believe some claims more strongly than others. Not just in that we feel more or less confident—though this often coincides with "believing more strongly" in our intended sense—but rather in that we judge some claims more likely. These likelihood judgments are known variously as graded beliefs, partial beliefs, degrees of belief, degrees of confidence, and credences. The existence of credences is a psychological fact; however they come about, credences are an integral component of many individual's cognition.

Credences have a number of obvious features. First, they are possessed by agents. Second, different agents are liable to possess different, even radically different, credences. Subject to the full breadth of human malfunction, we ought not expect any appreciable uniformity between agents.[2] Third and finally, credences are identical to neither feelings of confidence nor betting dispositions. Imagine, for instance, a Zen monk who lacks both any emotional response and any betting dispositions with respect to the claim 'the sun will rise tomorrow' (Eriksson and Hájek 2007). Such an individual may nevertheless judge this claim to be highly likely. As a result, while different credences are often associated with different emotional responses or betting dispositions, this association is not necessary.[3] Confidence and a willingness to bet at steep odds are hallmarks of high credence in a claim; they are not constitutive of it.

Where credences are the likelihood judgments of actual agents, rational credences are the likelihood judgments of ideally rational agents. Rational credences are thus perfected or idealized credences; the result of eliminating human fallibility from likelihood judgments. Unlike credences, all rational credences agree on at least some likelihood judgments. It would be irrational, for example, to have higher credence in a contradiction than a tautology or, more generally, to judge φ more likely than ψ when φ entails ψ. It is the task of accounts

[2]Actual agents' likelihood judgments fail both to range over propositions and to satisfy any meaningful formal criteria. On the former count, actual people routinely fail to recognize two statements as expressing the same proposition when they in fact do. Least contentiously, this occurs with simple calculation failures, e.g. someone versed in propositional logic who fails to notice the connection between "$\neg P$" and "$P \rightarrow \bot$". On the second count, it is certainly possible that a particular population exhibit nontrivial regularities. Subject to the full breadth of human malfunction, however, there are no universal regularities. Actual human judgments include not only failures of calculation and comprehension but also all those assessments made in a compromised state. Sleep deprivation, intoxication, and severe physical trauma suffice to undermine even those restrictions we might otherwise wish to declare universal truths of human judgment. There is nothing a suitably compromised individual cannot honestly reject.

[3]This example does not rule out a necessary relationship between credences and counterfactual betting dispositions, a position suggested by Ramsey (1931) and recently defended by Elliott (2019). Despite these endorsements, I find counterfactual betting dispositions more mysterious than a brute appeal to likelihood judgments, and so I retain the latter as a preferred gloss.

of rational credence to provide a full analysis of and justification for this common structure.

Orthodox accounts of rational credence uniformly hold that rational credences are, at minimum, probability functions. In these accounts, the state space Ω is interpreted as a set of possible worlds and the event space $\mathcal{F}$ as specifying all those propositions which admit of credal values. The probability function μ then represents a rational credence over $\mathcal{F}$, assigning greater or lesser values in $[0, 1]$ according to whether the proposition in question is judged to be more or less likely.

1.3 Confirmation

While rational credence is an idealization of a psychological phenomenon, confirmation arises as a generalization of logical consequence. Given a set of premises Σ, there has amassed a considerable consensus on what follows from this set, on what must be true given only Σ. This relationship between Σ and its consequences is popularly dubbed 'logical consequence' and straightforwardly underpins much of our reasoning. Despite these successes, logical consequence does not extend as far as desired. In point of fact, we are often confronted with claims which are neither entailed by nor inconsistent with our premises. Examples abound, but scientific theories—the theory of evolution, Newtonian mechanics, Mendelian genetics—and conspiracy theories—flat earth, fake moon landing, governments controlled by the Illuminati—are the most striking. Such theories are not treated as a single homogeneous whole. Instead, some are held to be better confirmed or better supported by the evidence.

As Karl Popper observed, there are actually two different relations we might plausibly call 'confirmation' here. The first is absolute confirmation or support of a claim by evidence. This is a relation between a set of evidence E and a single claim P or theory Γ.[4] This notion was formalized by Carnap (1962), for example, as $c(P|E)$ for c an appropriate probability function. A claim P was then more or less (absolutely) confirmed according to whether the probability assigned to P was closer or further from one. In contrast, incremental confirmation (sometimes also 'relative confirmation', 'relevance confirmation', or 'increase in firmness') is the relative support offered by a particular piece of evidence for a particular claim. This is a relation between a set of background evidence E, a single additional piece of evidence e, and a claim P. Using Carnap's theory again as an example, incremental confirmation was formalized not as the absolute probability of P relative to E

[4] This presentation is somewhat contentious. While Weisberg (2011) gives the same definition, Hájek and Joyce (2008) and Crupi and Tentori (2016) present absolute confirmation as a three-place relation between a set of background evidence E, a particular additional piece of evidence e, and a claim P. On their scheme, absolute confirmation of P by e (relative to background evidence E) refers to the probability of P given e (relative again to background evidence E). The separation of a evidential claim e from the main body of evidence only for it to be immediately rolled back into this greater collection is unnatural. The only apparent motivation here is to enforce an identical structure on both relations and to mimic the subjective Bayesian formalism (which completely suppresses mention of background evidence into the prior probability function). Maher (1996) presents a still different definition on behalf of Carnap, giving 'absolute confirmation' as increase in probability relative to no background evidence.

but rather as the incremental change induced by the addition of e: $c(P|E \cup \{e\}) - c(P|E)$. Intuitively, P is more or less (incrementally) confirmed by e according to how great or small the change in probability produced by e is. This dissertation is concerned with only the absolute confirmation relation; 'confirmation' simpliciter from here forward.

In practice, the confirmation relation is often presented under the heading of 'logical probability' (Gillies 2000; Hacking 2006; Weisberg 2011; Hájek 2012).[5] There are two good reasons to support this alternate title and one decisive point against it. First, 'logical probability' was the name of choice in the most influential entries on the topic: Keynes (1921) and Carnap (1962). Second, 'logical probability' is an apt description of the project itself so long as 'probability' is understood as an informal term for likelihood or weight. As Keynes (1921) explains,

> We are claiming, in fact, to cognise correctly a logical connection between one set [of] propositions which we call our evidence and which we suppose ourselves to know, and another set which we call our conclusions, and to which we attach more or less weight according to the grounds supplied by the first. . . . It is not straining the use of words to speak of this as the relation of probability. (5-6)

Despite these good reasons, the increasing shift towards a formal definition of 'probability' over the course of the last century makes 'logical probability' a misleading label today. No reason has been provided for thinking that confirmation is productively analyzed in terms of the mathematical theory of probability. Indeed, an unreflective slide into the mathematical theory of probability is precisely where previous accounts of confirmation go wrong.

As originally understood, the confirmation relation is both objective and normative. The relation is objective in that it is independent of any person or persons. Confirmation does not shift based on the individual considered, nor does it depend on our biology,[6] nor does it suppose any particular conception or perception of the world. Just like logical consequence, confirmation is an *a priori* relation to be discovered rather than invented. The confirmation relation is normative in that, absent either additional evidence or restrictions, it ought to guide both our internal convictions and our decision making. Confirmation describes the best that can be done on both epistemic and practical grounds in an ideal setting.

[5]Following Carnap (1962), confirmation is also sometimes described as 'inductive probability'. While Carnap (1962) takes validating induction as one of his desideratum, this does not follow from any of the glosses provided here. Indeed, I regard it as implausible that a relation of a kind with logical consequence should validate induction in the absence of further assumptions.

[6]Keynes (1921) denies this and is rightly criticized for this choice by Ramsey (1931). Keynes' hesitance on this point derives from the alleged possibility of logical relations which humanity is incapable of recognizing. As the eventual discussion of classical logic makes clear, no such worries attach to the view of logical consequence advocated here.

1.4 Outline

The remaining four chapters of this dissertation are divided into two parts. The first part focuses on the structure of rational credence and the inadequacy of probabilistic accounts thereof. The second chapter introduces the topic with a discussion of two different formalisms for ranking objects. The rest of the chapter then applies both formalisms to the case of rational credence with an emphasis on characterizing (i) those constraints that follow directly from logic and (ii) the consequences of supposing uniform combination of credal values over inconsistent sentences. The work of this chapter is largely programmatic in nature and reappears sporadically throughout the rest of the dissertation.

The third chapter leverages this initial discussion into a criticism of probabilistic accounts of rational credence. Contemporary discussion of rational credence is dominated by probabilism in general and subjective Bayesianism in particular. The former holds that rational credences are probability functions and the latter that this is the only synchronic constraint on rational credence. Naturally, a large number of arguments have been offered in defense of both. This chapter surveys three of the most influential justifications for probabilism: Dutch book arguments, representation arguments, and gradational accuracy arguments. In all three cases, the proposed argument is demonstrated to presuppose both comparability and the Archimedean property, key tenets of probabilism. As a direct result, these orthodox accounts of rational credence fail to justify their identification of rational credences with probability functions.

The fourth chapter begins the dissertation's second part: a non-probabilistic account of confirmation which is faithful to the intuitions sketched above, viz. a degreed extension of classical logic which is both objective and normative. The contemporary case against confirmation provides three immediate barriers: Ramsey's skepticism, d'Alembert's riddle, and Bertrand's paradox. The fourth chapter considers the first two of these objections and constructs a finite account of confirmation that overcomes both.

The fifth and final chapter begins with an extended commentary on the most substantial of the three objections to confirmation: Bertrand's paradox. Both standard characterizations of and replies to Bertrand's paradox are shown to underestimate the underlying phenomenon. The first contribution of the chapter is thus a revised characterization of Bertrand's paradox that significantly expands its scope. Despite this generalization, the paradox succeeds only in blocking probabilistic accounts of confirmation. The second contribution of this chapter is then an extension of the finite account of confirmation from chapter four which successfully evades the paradox. Further, it is shown that this account of confirmation is a maximal solution to the paradox; any alternative account of confirmation either falls victim to the paradox or is strictly weaker. This chapter ends with a critical discussion of degrees of confirmation and the extent to which the proposed account supports this talk.

The work of both parts of this dissertation can be brought together with a pair of observations. First, any adequate account of confirmation is prima facie an adequate account of rational credence. What more could one want from a theory of rational credence than an objective and normative relation representing the evidential support for a claim? Second, among the characteristic features of the proposed account of confirmation are failures of both comparability and the Archimedean property, the very same properties left unjustified by contemporary accounts of rational credence. The negative work of part one thus dovetails with the positive account offered in part two; confirmation is no misleading misconception, and rational credences are not probability functions.

Part I

Formalizing Rational Credence

Chapter 2

Likelihood Spaces and Structures

This chapter explores the task of formalizing rational credence with an emphasis on identifying minimal constraints rather than advocating any particular account. The first section presents and compares two different formalisms for ranking objects: comparative rankings and absolute rankings. The second section applies the latter to the case of rational credence, characterizing both the constraints that follow directly from logic and a natural extension which supposes uniformity in combining inconsistent sentences. The third and final section is largely technical in nature, establishing simple correspondences between the work on absolute rankings in the second section and comparative rankings.

2.1 Comparative and Absolute Rankings

In ranking any class of objects, there are two formally distinct approaches available. First, objects may be placed into a relative or comparative ranking; a theory may stipulate when an object o is above o', when o and o' are equal, and when o' is at least o. The following abbreviations for these relationships are typical:

$o \prec o'$	o is less than o'.
$o \sim o'$	o and o' are equal.
$o \precsim o'$	o is less than or equal to o'.

Second, objects may be ranked via association with some absolute scale $\langle A, \leq, < \rangle$. That is, a theory may stipulate a function $\mu : O \rightarrow A$ which assigns each object $o \in O$ a value $a \in A$. Particular objects o are then more/less/equal according to the position of their value $\mu(o)$ in the absolute scale.

Formally, the presentation above overspecifies both approaches. Given only a binary relation $\precsim$ on O, a corresponding strict relation $\prec$ and symmetric relation $\sim$ can be defined by

(i) $o \prec o'$ if and only if $o \precsim o'$ and $o' \not\precsim o$.

(ii) $o \sim o'$ if and only if $o \precsim o'$ and $o' \precsim o$.

Similarly, given that $\sim$ is symmetric:

If $o \sim o'$, then $o' \sim o$.

and $\prec$ is exclusive:

If $o \prec o'$, then both $o' \nprec o$ and $o \nsim o'$.

a corresponding non-strict relation $\precsim$ can be defined by

(iii) $o \precsim o'$ if and only if $o \prec o'$ or $o \sim o'$.

Proposition 2.1

Let O be a set and each of $\precsim, \prec, \sim$ binary relations over O. Then, the following are equivalent:

(1) (i) and (ii) hold;

(2) (iii) holds, $\sim$ is symmetric, and $\prec$ is exclusive.

Accepting (i)-(iii), it thus suffices to supply either a non-strict comparative relation $\precsim$ or the pair $\langle \prec, \sim \rangle$ where $\sim$ is symmetric and $\prec$ is exclusive.

Parallel observations apply to absolute rankings. The comparative $\sim$ relation here is induced by the identity relation over A and thus satisfies symmetry. Supposing (i)-(iii) and that the strict relation $<$ satisfies exclusivity, an absolute scale may be taken to be either a set and non-strict relation $\langle A, \leq \rangle$ or a set and strict relation $\langle A, < \rangle$ with the missing relation derived by either (i) or (iii). The choice then between $\langle A, \leq \rangle$ and $\langle A, < \rangle$ is likewise of little formal consequence.

For simplicity, the convention of using the non-strict formulation of both approaches is adopted here.

Definition 2.1 A *comparative ranking* over a set of objects O is a binary relation $\precsim$ over O satisfying

- *Reflexivity.* For any $o \in O$, $o \precsim o$.
- *Transitivity.* For any $o, o', o'' \in O$, if $o \precsim o'$ and $o' \precsim o''$, then $o \precsim o''$.

Since the values in an absolute scale measure different sizes, equality of size occurs only with identity between values:

Definition 2.2 An *absolute scale* $\langle A, \leq \rangle$ is a ordered pair of a set and a binary relation $\leq$ over A satisfying

- *Reflexivity.* For any $a \in A$, $a \leq a$.
- *Transitivity.* For any $a, a', a'' \in A$, if $a \leq a'$ and $a' \leq a''$, then $a \leq a''$.
- *Antisymmetry.* For any $a, a' \in A$, if $a \leq a'$ and $a' \leq a$, then $a = a'$.

Definition 2.3 An *absolute assignment* over a set of objects O using absolute scale $\langle A, \leq \rangle$ is a function $\mu : O \to A$.

Comparative rankings are thus preorders while absolute assignments assign values from a partial order.

Without further restrictions, there is no representational difference between comparative and absolute rankings. Given a theory which produces an absolute ranking μ, a corresponding comparative theory can be had simply by defining

$$o \precsim_{\mu} o' \Leftrightarrow \mu(o) \leq \mu(o').$$

Similarly, a theory which produces a comparative ranking can be replicated in an absolute framework by taking equivalence classes under the comparative $\sim$ relation as an absolute scale for O.

Proposition 2.2

Let a set of objects O be given. Then,

(i) For any comparative ranking $\precsim$ over O, there exists an absolute scale $\langle A, \leq \rangle$ and absolute ranking $\mu : O \to A$ such that

$$o \precsim o' \Leftrightarrow \mu(o) \leq \mu(o').$$

(ii) For any absolute scale $\langle A, \leq \rangle$ and absolute ranking $\mu : O \to A$, there exists a comparative ranking $\precsim$ over O such that

$$o \precsim o' \Leftrightarrow \mu(o) \leq \mu(o').$$

In the most general case, absolute and comparative rankings are entirely interchangeable over a fixed object set O.

This straightforward relationship can, however, become significantly more complex with the addition of further constraints. It is useful here to recognize three broad categories. First, constraints may be imposed on the comparative ranking. It might be required, for instance, that the ranking be total or that a particular operation on objects interact nicely with the comparative ranking. Second, constraints may be imposed on either the absolute scale $\langle A, \leq \rangle$ or on the absolute assignment $\mu : O \to A$. Work in formal epistemology, for example, often supposes that only absolute assignments using the absolute scale $\langle [0, 1], \leq \rangle$ are allowed. With constraints in either of these two categories, maintaining agreement between absolute assignments and comparative rankings necessitates the introduction of parallel restrictions on the other formalism. Identifying both adequate and intuitive choices here can be a matter of considerable difficulty.

The final category of constraints concerns not comparisons over a particular set of objects O but rather cross-set comparisons. In many applications, a homogeneous ranking is desired over a collection of object sets $\mathcal{O} = \{O, \ldots\}$. If this collection $\mathcal{O}$ is not closed under union, novel cross-set commitments may arise. Such commitments are particularly natural with absolute rankings. Taking, for example, two absolute assignments μ and μ' over two disjoint object sets O and O' which share an absolute scale $\langle A, \leq \rangle$, it is natural (though not necessary) to interpret μ and μ' as fixing not only a comparative ranking $\precsim$ over O and a comparative ranking $\precsim'$ over O' but also the comparative ranking $\precsim^+$ over $O \cup O'$ defined by

$$o_1 \precsim^+ o_2 \text{ if and only if } \mu^+(o_1) \leq \mu^+(o_2)$$

where

$$\mu^+(o) = \begin{cases} \mu(o) & \text{if } o \in O \\ \mu'(o) & \text{if } o \in O'. \end{cases}$$

In general, endorsing cross-set comparisons may again introduce a more complicated relationship between comparative and absolute rankings.

2.2 Likelihood Spaces and Structures

Providing a formal characterization of rational credence in terms of absolute rankings requires three distinct components. First is the specification of those entities to which credal values will attach. Second is the choice of an absolute scale $\langle \mathbb{P}, \leq \rangle$ for these values. Third and finally, we may concern ourselves with interactions between the first and second components—how it is that values from the endorsed absolute scale may be assigned. It is not difficult to see that there are dependencies between these three tasks—one should not, for instance, be in the business of assigning values which do not exist in one's scale. Conversely, however, each task also possesses a significant degree of independence from the others; a space of values does not generate the entities to which these values may be applied, nor does a space of values and a set of entities necessitate particular interactions between them. Each individual step above involves a non-trivial choice by the theorist.

2.2.1 The Objects of Rational Credence

Conventional accounts of rational credence almost exclusively assign credal values to sets of possibilities ("events" or "propositions") from some distinguished collection, cf. Halpern (2017). While this phrasing is evocative, it is nevertheless unclear what precisely a possibility (or possible world for that matter) is. Formal logic not only provides a ready replacement for this loose talk but one which helps fix a minimal structure for theories of rational credence. Rational credences properly apply to formal languages L which are equipped with a non-empty, exhaustive collection[1] of models or model space $\Omega_{\mathcal{L}}$.[2] A rational credence may then be taken to assign values either to the sentences of such a formal language or to subcollections of the model space $\Omega_{\mathcal{L}}$.

The interchange between sentences and collections of models is straightforward, and so we will shift between both perspectives. Given a formal language L and a model space $\Omega_{\mathcal{L}}$, every sentence $\varphi \in L$ is associated with the collection of models from $\Omega_{\mathcal{L}}$ which make it true, notated $[\![\varphi]\!]$. Sentences of L thus correspond to unique collections of models though a particular collection of models may correspond to many or even no sentences. As an example of this first, note that each of φ, $(\varphi \vee \varphi)$, and $((\varphi \vee \varphi) \vee \varphi)$ represent distinct sentences but all three correspond to the same collection of models $[\![\varphi]\!]$. In general, the collection of models formulation collapses syntactic distinctions which do not reflect any semantic difference. For the second, a collection of models corresponding to no sentence of the formal language L is the hallmark of an expressive weakness, a meaningful but inexpressible "proposition". While some care must thus be taken in the transition between sentences and collections of models, the difficulties here are outweighed by the utility—both philosophical and mathematical—of the dual perspective.

We will suppose both that all of the formal languages L at issue are equipped with Boolean connectives and that any theory of rational credences applies at minimum to *propositional languages*:

[1] Collections (sometimes also, classes) are a generalization of the notion of set. Every set is a collection but not every collection is a set, e.g., the collection of all first-order models. I assume throughout that the collections considered satisfy the axioms of ZFC.

[2] What of formal languages—e.g., propositional modal languages—which do not possess an accepted, exhaustive collection? Such languages necessarily lack a fixed notion of model-theoretic consequence which—all things considered—is a larger and more immediate shortcoming than a breakdown in the notion of rational credence. We may proceed by supposing that a particular collection of models is exhaustive though this should be recognized as an additional and often substantial assumption.

Definition 2.4 Given a set of propositional letters σ, the *propositional language L generated from σ* is the set containing all and only

(i) $\bot$,

(ii) $\top$,

(iii) P_i for every $P_i \in \sigma$,

(iv) For any $\varphi, \psi \in L$,

$$\neg\varphi \ | \ (\varphi \wedge \psi) \ | \ (\varphi \vee \psi) \ | \ (\varphi \rightarrow \psi) \ | \ (\varphi \leftrightarrow \psi).$$

While it is possible to use both simpler and more complex languages, the propositional setting is both familiar and uncontroversial. However far theories of rational credence extend, they cannot fail to apply to propositional languages.

Every propositional language is associated with a straightforward and widely-accepted model space.[3] The relevant definitions are briefly reproduced below:

Definition 2.5 For a set of propositional letters σ and associated propositional language L, an *L-model* is a function $v : \sigma \rightarrow \{T, F\}$. Given an L-model M, truth in M is defined recursively by:

(i) $M \Vdash P_i$ if and only if $v(P_i) = T$;

(ii) $M \Vdash \top$;

(iii) $M \nVdash \bot$;

(iv) For $\varphi, \psi \in L$,

$$\begin{aligned}
M \Vdash \neg\varphi &\text{ if and only if } M \nVdash \varphi \\
M \Vdash (\varphi \wedge \psi) &\text{ if and only if } M \Vdash \varphi \text{ and } M \Vdash \psi \\
M \Vdash (\varphi \vee \psi) &\text{ if and only if } M \Vdash \varphi \text{ or } M \Vdash \psi \\
M \Vdash (\varphi \rightarrow \psi) &\text{ if and only if } M \nVdash \varphi \text{ or } M \Vdash \psi \\
M \Vdash (\varphi \leftrightarrow \psi) &\text{ if and only if either } M \Vdash \varphi \text{ and } M \Vdash \psi \text{ or } M \nVdash \varphi \text{ and } M \nVdash \psi.
\end{aligned}$$

[3] Other choices—e.g., a model space allowing differences in signature or a model space which omits particular models—may nevertheless be useful depending on context

Definition 2.6 For $\Gamma \subseteq L$ and $\varphi \in L$, φ *is a consequence of* Γ, notated $\Gamma \models \varphi$, if and only if for every L-model M, if $M \Vdash \psi$ for every $\psi \in \Gamma$, then $M \Vdash \varphi$ as well.

- A sentence $\varphi \in L$ is a *contradiction* if and only if for every L-model M, $M \nVdash \varphi$.
- A sentence $\varphi \in L$ is *valid* if and only if for every L-model M, $M \Vdash \varphi$.
- A set of sentences $\Gamma \subseteq L$ is *consistent* if and only if there exists an L-model M such that $M \Vdash \psi$ for every $\psi \in \Gamma$.

We require that every model in a model space $\Omega_{\mathcal{L}}$ for L assigns a truth value to every sentence in L. For any $\varphi \in L$, the corresponding set of models in $\Omega_{\mathcal{L}}$ is thus given by

$$[\![\varphi]\!] = \{v \in \Omega_{\mathcal{L}} \ : \ v \Vdash \varphi\}.$$

Indeed, since L is equipped with Boolean connectives, the set of models formulation naturally gives rise to the algebra[4]

$$\mathcal{F}_{\mathcal{L}} = \{[\![\varphi]\!] \ : \ \varphi \in \mathcal{L}\}.$$

The typical emphasis on sets of possibilities and algebras in theories of rational credence thus falls out of the current approach.

In sum, rational credences are provided relative to a formal language L and model space $\Omega_{\mathcal{L}}$. With a little care, credal values may then be attributed either to sentences of L or to subcollections of the model space $\Omega_{\mathcal{L}}$. Given our insistence on Boolean connectives and supposing that $\Omega_{\mathcal{L}}$ is a set, these subcollections of $\Omega_{\mathcal{L}}$ form an algebra $\mathcal{F}_{\mathcal{L}}$ over $\Omega_{\mathcal{L}}$. The most obvious examples of formal languages equipped with model spaces are particular propositional languages together with the set of all their models. Finally, we will require that theories of rational credence apply to propositional languages and their associated model spaces at a minimum.

2.2.2 The Space of Values

Turning to the second component of a theory of rational credence, an absolute scale for rational credal values is a partially ordered set with two distinguished extreme elements. Such a scale will be called a likelihood space.

[4]An *algebra of sets* is a collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ which is closed under finite union, finite intersection, and complementation.

Definition 2.7 A *likelihood space* $\mathbb{P}$ is a four-tuple containing

(i) a domain set, $\mathrm{dom}(\mathbb{P})$,

(ii) a distinguished impossible element, $\dot{0} \in \mathrm{dom}(\mathbb{P})$,

(iii) a distinguished certain element, $\dot{1} \in \mathrm{dom}(\mathbb{P})$,

(iv) a binary relation $\preceq$ on $\mathrm{dom}(\mathbb{P})$ satisfying

   (a) *Non-triviality*
   $\dot{0} \prec \dot{1}$.

   (b) *Transitivity*
   For any $p_1, p_2, p_3 \in \mathrm{dom}(\mathbb{P})$, if $p_1 \preceq p_2$ and $p_2 \preceq p_3$, then $p_1 \preceq p_3$.

   (c) *Reflexivity*
   For any $p \in \mathrm{dom}(\mathbb{P})$, $p \preceq p$.

   (d) *Anti-Symmetry*
   For any $p_1, p_2 \in \mathrm{dom}(\mathbb{P})$, if $p_1 \preceq p_2$ and $p_2 \preceq p_1$, then $p_1 = p_2$.

   (e) *Boundedness*
   For any $p \in \mathrm{dom}(\mathbb{P}) - \{\dot{0}, \dot{1}\}$, $\dot{0} \prec p \prec \dot{1}$.

The only novelty here over and above absolute scales generally is the introduction of a distinct lower and upper bound. These two elements correspond to impossibility and certainty respectively and reflect a necessary feature of any adequate theory of rational credence. Whatever credal values are, they are necessarily partially ordered and range between impossibility and certainty.

Convention 1. Set-theoretic operations applied to likelihood spaces are understood as operations on the domain of the space, e.g., $p \in \mathbb{P}$ will be used instead of $p \in \mathrm{dom}(\mathbb{P})$.

Convention 2. Likelihood spaces $\mathbb{P}$ will often be abbreviated to a pair of a domain and a binary relation, $\langle \mathrm{dom}(\mathbb{P}), \preceq \rangle$ with the choice of $\dot{0}$ and $\dot{1}$ implicit as the smallest and largest elements in the domain according to $\preceq$.

The actual representatives within a specified domain $\mathrm{dom}(\mathbb{P})$ are merely conventional; it is the structural features of the space $\mathbb{P}$ which are of interest. Accordingly,

Definition 2.8 A likelihood space $\mathbb{P}_1$ is *equivalent* to a likelihood space $\mathbb{P}_2$ if and only if there exists a $\preceq$-preserving bijection $\delta : \mathbb{P}_1 \to \mathbb{P}_2$.

For those familiar with modern formal logic, this may be equivalently stated as $\mathbb{P}_1$ and $\mathbb{P}_2$ are isomorphic when interpreted as first-order models of the language generated from the signature $\langle \dot{0}, \dot{1}, \leq \rangle$. Under this criteria, each of the following likelihood spaces is distinct:

$$\langle\ \{0,1\}, 0, 1, \leq\ \rangle$$

$$\langle \{0, \frac{1}{2}, 1\}, 0, 1, \leq\ \rangle$$

$$\langle [0,1], 0, 1, \leq\ \rangle$$

$$\langle [0,1] \cup \{\frac{1}{\infty}\} \cup \{\frac{\infty - 1}{\infty}\}, 0, 1, \leq\ \rangle$$

Each of these spaces naturally suggests a position in philosophical discussion of rational credence. The first is naturally construed as representing a complete lack of epistemic uncertainty, the second as a view which admits of only a degree-less uncertainty, the third as the conventional probability account, and the last as a simplistic infinitesimal account.

This slide from likelihood space to a full theory of rational credence is—however—to be resisted. A likelihood space alone does not force an account of what these likelihoods attach to, nor—more critically—how these probabilities vary relative to their bearers. To take a salient example, Jeffreys (1967) entertains the space $\langle [0, \infty], 0, \infty, \leq \rangle$ for situations which contain a countably infinite number of disjoint events. It is not difficult to show, moreover, that this likelihood space is equivalent to $\langle [0, 1], 0, 1, \leq \rangle$ in our intended sense. Jeffreys is not thereby offering, however, the conventional account in disguise. $\leq$-preserving bijection between likelihood spaces takes no account of the behavior of likelihood values *relative to a language and model space*, and it is here that the two proposals differ.

Identifying disjunction between incompatible alternatives with addition on $[0, \infty]$, the induced disjunction operation on the space $\langle [0, 1], 0, 1, \leq \rangle$ necessarily possesses instances of sub-additivity. That is, for any $\leq$-preserving bijection $\delta : [0, \infty] \rightarrow [0, 1]$, there exists $p, p' \in [0, \infty]$ such that $\delta(p + p') < \delta(p) + \delta(p')$. Conversely, identifying disjunction between incompatible alternatives with the standard addition relation over $[0, 1]$, the induced disjunction operation on $\langle [0, \infty], 0, 1, \leq \rangle$ has instances of super-additivity. In each case, the association of the disjunction of disjoint alternatives with addition over the likelihood space represents an additional and substantial commitment, one which renders the two interpretations incompatible despite the equivalence of the likelihood spaces themselves. To reiterate the caution which headed this chapter, the identification of a likelihood space does not—despite conventional associations—by itself settle either the objects of rational credence or interactions between these objects and the values assigned.

2.2.3 Interactions

Third and finally, a theory of rational credence must determine how values from the likelihood space may be assigned. Having taken formal languages equipped with a model space as a basis for theories of rational credence, this component requires identifying some collection of functions from the language L (or algebra $\mathcal{F}_{\mathcal{L}}$) to the likelihood space $\langle \mathbb{P}, \leq \rangle$ as all and only the rational credences. This is, for example, the role of Kolmogorov's axioms in the subjective Bayesian account of rational credence. Historically, this third component has proven the most contentious of the three.

A number of restrictions on likelihood assignments follow from the requirement that rational credences ought to refine logic. These restrictions may be summarized in three principles:

N. *Normativity of Logic*
For any rational credence μ and any $\varphi, \psi \in L$, if $\varphi \models \psi$, then $\mu(\varphi) \preceq \mu(\psi)$.

C. *Certain Validities*
For any rational credence μ and any $\varphi \in L$, if $\models \varphi$, then $\mu(\varphi) = \dot{1}$.

I. *Impossible Contradictions*
For any rational credence μ and any $\varphi \in L$, if $\models \neg\varphi$, then $\mu(\varphi) = \dot{0}$.

The model-theoretic consequence relation $\models$ is here defined as usual, viz. given L and $\Omega_{\mathcal{L}}$, every model in $\Omega_{\mathcal{L}}$ that makes φ true also makes ψ true. A formal language L and model space $\Omega_{\mathcal{L}}$ thus suffice for a corresponding notion of consequence, a logic for L together with $\Omega_{\mathcal{L}}$. In terms of collections of models and the algebra $\mathcal{F}_{\mathcal{L}}$, these three principles require that likelihoods do not increase when moving to subcollections, that $\Omega_{\mathcal{L}}$ receives the value $\dot{1}$, and that the empty set receives the value $\dot{0}$.[5] Put together, *Normativity of Logic*, *Certain Validities*, and *Impossible Contradictions* require only that rational credences exist within the bounds of a language's own logic.

In the case of a propositional language L, *Normativity of Logic* thus guarantees, for example, that $\mu(P) \preceq \mu((P \vee Q))$ and $\mu((P \wedge Q)) \preceq \mu(P)$ for any rational credence μ. *Certain Validities* meanwhile requires that $\mu(\top) = \mu((P \vee \neg P)) = \dot{1}$ and, more generally, that all valid sentences receive maximal credence. Finally, *Impossible Contradictions* requires that all contradictions receive minimal credence. Thus, $\mu(\bot) = \mu((P \wedge \neg P)) = \dot{0}$ for any rational credence μ.

Not all restrictions on rational credences need reduce to logic. Indeed, even relatively permissive contemporary accounts embrace some extra-logical commitments. Drawing on Krantz et al. (1971), many of these additional commitments can be usefully summarized by stipulating that a particular operation on either sentences or collections of models corresponds to a particular operation on likelihood values. In the case of probability, for example, a correspondence between disjunction of inconsistent sentences (equivalently, unions

[5] A similar framework was independently developed by Friedman and Halpern (1995); Halpern (2001, 2017) under the heading of 'plausibility measures'. A *plausibility space* is a triple $\langle W, \mathcal{F}, \mu \rangle$ where W is a set of possibilities, $\mathcal{F}$ is an algebra over W, and μ is a function from $\mathcal{F}$ to a partially ordered set with two extreme elements which satisfies the set-theoretic equivalents of (N), (C), and (I). It is not difficult to prove that every formal language L, model space $\Omega_{\mathcal{L}}$, and function $\mu : L \to \mathbb{P}$ for a likelihood space $\langle \mathbb{P}, \preceq \rangle$ which satisfies the logical constraints (N), (C), and (I) gives rise to a corresponding plausibility space. The explicit use of formal languages and model spaces is specific to the current approach and casts Halpern's plausibility measure formalism in a new light. Likelihood assignments satisfying (N), (C), and (I) are not simply one more general formalism among many proposals; these are the weakest constraints consistent with refining a language's logic. A few of Friedman and Halpern (1995)'s results will prove useful in the next section though the eventual positive proposal deviates radically.

of disjoint sets of models) and addition is typically postulated. After all, if φ and ψ are inconsistent, the likelihood of $(\varphi \vee \psi)$ ought to just be the result of 'combining' the likelihood of φ with the likelihood of ψ. This particular approach is both ubiquitous and intuitive enough that it will be useful to have a shorthand for a likelihood space equipped with an operation which corresponds to the disjunction of inconsistent sentences.

While actually stipulating the use of addition would beg the question in favor of particular likelihood spaces, an inconsistent disjunction operation ought to be at least a partial function[6] from $\mathbb{P} \times \mathbb{P}$ to $\mathbb{P}$ and addition-like:

Definition 2.9 A *likelihood structure* $\langle \mathbb{P}, \circ \rangle$ is a pair containing

(i) a likelihood space $\mathbb{P}$,

(ii) an operation $\circ : \mathbb{P} \times \mathbb{P} \twoheadrightarrow \mathbb{P}$ satisfying

  (a) *Commutativity*
  For any $p_1, p_2 \in \mathbb{P}$, if $\circ(p_1, p_2)\downarrow$, then $\circ(p_2, p_1)\downarrow = \circ(p_1, p_2)$.

  (b) *Additive Identity*
  For any $p \in \mathbb{P}$, $\circ(p, \dot{0})\downarrow = p$.

  (c) *Existence of Complements*
  For any $p \in \mathbb{P}$, there exists p' such that $\circ(p, p')\downarrow = \dot{1}$.

  (d) *Associativity*
  For any $p_1, p_2, p_3 \in \mathbb{P}$, if both $\circ(p_1, \circ(p_2, p_3))\downarrow$ and $\circ(p_1, \circ(p_2, p_3))\downarrow$, then $\circ(\circ(p_1, p_2), p_3) = \circ(p_1, \circ(p_2, p_3))$.

  (e) *Monotonicity*
  For any $p, p_1, p_2 \in \mathbb{P}$ with $\circ(p, p_1)\downarrow$ and $\circ(p, p_2)\downarrow$, $p_1 \preceq p_2$ if and only if $\circ(p, p_1) \preceq \circ(p, p_2)$.

While intuitive, this simple move from likelihood spaces to likelihood structures enforces three further non-logical restrictions on likelihood assignments.

Convention 3. Likelihood structures $\langle \mathbb{P}, \circ \rangle$ will often be presented as a triple of a domain dom($\mathbb{P}$), a binary relation $\preceq$, and an operation $\circ : \mathbb{P} \times \mathbb{P} \twoheadrightarrow \mathbb{P}$ with the choice of $\dot{0}$ and $\dot{1}$ implicit as the smallest and largest elements in the domain according to $\preceq$.

First, an arbitrary likelihood assignment $\mu : L \to \mathbb{P}$ satisfying the logical constraints (N), (C), and (I) need not induce a well-defined operation for the disjunction of inconsistent sentences. Let L be the propositional language generated from $\sigma = \{P, Q\}$ and $\Omega_{\mathcal{L}}$ the set $\{v_1, v_2, v_3, v_4\}$ of all propositional models for L. Consider the likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to [0, 1]$ defined by

[6] A *partial function* $f : A \twoheadrightarrow B$ for two sets A and B is a relation on $A \times B$ such that for any $a \in A$, there is either exactly one $b \in B$ with $\langle a, b \rangle \in f$ or no $b \in B$ with $\langle a, b \rangle \in f$. In the former case, we say "f is defined on input a" and write $f(a)\downarrow = b$ or $f(a)\downarrow$. In the latter, we say "f is undefined on input a" and write $f(a)\uparrow$.

$$\mu(\{v_1, v_2, v_3, v_4\}) = \dot{1}$$

$$\vee$$

$$\mu(\{v_1, v_2, v_3\}) = \mu(\{v_1, v_2, v4\}) = \mu(\{v_1, v_3, v_4\}) = \mu(\{v_2, v_3, v_4\}) = \frac{4}{5}$$

$$\vee$$

$$\mu(\{v_1, v_3\}) = \mu(\{v_1, v_4\}) = \mu(\{v_2, v_3\}) = \mu(\{v_3, v_4\}) = \mu(\{v_2, v_4\}) = \frac{3}{5}$$

$$\vee$$

$$\mu(\{v_1, v_2\}) = \frac{2}{5}$$

$$\vee$$

$$\mu(\{v_1\}) = \mu(\{v_2\}) = \mu(\{v_3\}) = \mu(\{v_4\}) = \frac{1}{5}$$

$$\vee$$

$$\mu(\emptyset) = \dot{0}.$$

Despite satisfying (N), (C), and (I), the likelihoods of unions over $\{v_1\}$, $\{v_2\}$, and $\{v_3\}$ depend not just on the likelihood assigned to each set individually but also on the particular set in question. The union of $\{v_1\}$ and $\{v_2\}$, for example, produces a set with likelihood strictly below the union of $\{v_2\}$ and $\{v_3\}$ despite $\{v_1\}$, $\{v_2\}$, and $\{v_3\}$ being assigned equal likelihood.

While intuitive, the introduction of an operation for the disjunction of inconsistent sentences thus introduces an additional functionality restriction on likelihood assignments.

F. *Functionality for Inconsistent Disjunction / Disjoint Union*
For any two pairs of inconsistent sentences φ, ψ and φ', ψ', if $\mu(\varphi) = \mu(\varphi')$ and $\mu(\psi) = \mu(\psi')$, then $\mu((\varphi \vee \psi)) = \mu((\varphi' \vee \psi'))$.

Given a likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ which satisfies (F), the induced operator $\circ$ for disjunction of inconsistent sentences satisfies the commutativity, additive identity, and existence of complements restrictions over the values used by μ, $\mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}$.

Proposition 2.3

Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$ satisfying (N), (C), and (I), μ satisfies (F) only if $\circ$ satisfies commutativity, additive identity, and existence of complements over $\mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}$ where $\circ : \mathbb{P} \times \mathbb{P} \twoheadrightarrow \mathbb{P}$ is defined by

$$\mu(\varphi) \circ \mu(\psi) = \mu((\varphi \vee \psi))$$

for $\varphi, \psi \in L$ inconsistent.

Absent from this result are the associativity and monotonicity restrictions imposed on the inconsistent disjunction operator $\circ$ in likelihood structures. Given a triple of inconsistent sentences φ, ψ, γ, any likelihood assignment $\mu : L \to \mathbb{P}$ satisfying the logical constraints (N), (C), and (I) as well as the non-logical constraint (F) is associative over φ, ψ, and γ, i.e.,

$$\mu(\varphi) \circ [\mu(\psi) \circ \mu(\gamma)] = [\mu(\varphi) \circ \mu(\psi)] \circ \mu(\gamma).$$

While this is sometimes taken to show that $\circ$ is associative in general—e.g., by Fine (1973)—this is not true. The following example shows that an arbitrary likelihood assignment $\mu : L \to \mathbb{P}$ satisfying (N), (C), (I), and (F) need not induce an operation for the disjunction of inconsistent sentences which satisfies associativity. The existence of such an example was first noted by Friedman and Halpern (1995) but was never explicitly presented. Let L be the propositional language generated from $\sigma = \{P, Q, R\}$ and $\Omega_{\mathcal{L}}$ a subset $\{v_1, v_{2a}, v_{2b}, v_{2c}, v_3, v_4\}$ of all propositional models for L. Using only the subscript for each model while dropping both μ and explicit set notation in order to reduce clutter,

$12_a2_b2_c34$

$2_a2_b2_c34$

$12_a2_b34 \sim 12_b2_c34 \sim 12_a2_c34$

$2_a2_b34 \sim 2_b2_c34 \sim 2_a2_c34$

12_a34

12_b34

12_c34

$2_a34 \sim 2_b34 \sim 2_c34 \sim 12_a2_b2_c4$

134

$2_a2_b2_c4$

$12_a2_b4 \sim 12_b2_c4 \sim 12_a2_c4 \sim 34$

$2_a2_b4 \sim 2_b2_c4 \sim 2_a2_c4$

$\mathbf{12_a4 \nsim 12_b4 \nsim 12_c4}$

$\mathbf{2_a4 \nsim 2_b4 \nsim 2_c4}$

14

4

$12_a2_b2_c3$

$2_a2_b2_c3$

$12_a2_b3 \sim 12_b2_c3 \sim 12_a2_c3$

$2_a2_b3 \sim 2_b2_c3 \sim 2_a2_c3$

$\mathbf{12_a3 \nsim 12_b3 \nsim 12_c3}$

$2_a3 \sim 2_b3 \sim 2_c3 \sim 12_a2_b2_c$

13

$2_a2_b2_c$

$12_a2_b \sim 12_b2_c \sim 12_a2_c \sim 3$

$2_a2_b \sim 2_b2_c \sim 2_a2_c$

$\mathbf{12_a \nsim 12_b \nsim 12_c}$

$\mathbf{2_a \nsim 2_b \nsim 2_c}$

1.

The $\prec$ relation is encoded here based on the vertical placement of sets while incomparabilities are both explicitly specified and bolded for clarity. The induced $\circ$ operation is not associative in this example since

$$\begin{aligned} \mu(12_b34) &\prec \mu(12_a34) \\ \mu(3) \circ \mu(12_b4) &\prec \mu(1) \circ \mu(2_a34) \\ \mu(12_a2_c) \circ \mu(12_b4) &\prec \mu(1) \circ \mu(12_a2_b2_c4) \\ [\mu(1) \circ \mu(2_a2_c)] \circ \mu(12_b4) &\prec \mu(1) \circ [\mu(2_a2_c) \circ \mu(12_b4)]. \end{aligned}$$

This failure relies critically on a scarcity of sets with particular sizes. If this last row were expressible in terms of a collection of pairwise disjoint sets, then (F) together with the logical restrictions entails associativity.

The second non-logical restriction built into likelihood structures blocks cases of precisely this sort. The induced inconsistent disjunction operator $\circ$ ought to be associative even if there are too few sets to force this property:

A. *Associativity for Inconsistent Disjunction / Disjoint Union*
For any φ, ψ, γ, if both $\mu(\varphi) \circ [\mu(\psi) \circ \mu(\gamma)] \downarrow$ and $[\mu(\varphi) \circ \mu(\psi)] \circ \mu(\gamma) \downarrow$, then $\mu(\varphi) \circ [\mu(\psi) \circ \mu(\gamma)] = \mu(\varphi) \circ [\mu(\psi) \circ \mu(\gamma)]$.

Likelihoods ought to be insensitive to the order in which they are combined.

A final example establishes that an arbitrary likelihood assignment $\mu : L \to \mathbb{P}$ satisfying the logical constraints (N), (C), and (I) as well as both (F) and (A) need not induce an operation for the disjunction of inconsistent sentences which satisfies monotonicity. Let L be the propositional language generated from $\sigma = \{P, Q\}$ and $\Omega_{\mathcal{L}}$ the set $\{v_1, v_2, v_3, v_4\}$ of all propositional models for L. Consider the likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to [0, 1]$ defined by

$$\mu(\{v_1, v_2, v_3, v_4\}) = \dot{1}$$

$$\curlyvee$$

$$\mu(\{v_1, v_2, v_3\}) = \mu(\{v_1, v_2, v4\}) = \mu(\{v_1, v_3, v_4\}) = \mu(\{v_2, v_3, v_4\}) = \frac{5}{6}$$

$$\curlyvee$$

$$\mu(\{v_1, v_2\}) = \mu(\{v_1, v_3\}) = \mu(\{v_1, v_4\}) = \frac{2}{3}$$

$$\curlyvee$$

$$\mu(\{v_2, v_3\}) = \mu(\{v_3, v_4\}) = \mu(\{v_2, v_4\}) = \frac{1}{2}$$

$$\curlyvee$$

$$\mu(\{v_2\}) = \mu(\{v_3\}) = \mu(\{v_4\}) = \frac{1}{3}$$

$$\curlyvee$$

$$\mu(\{v_1\}) = \frac{1}{6}$$

$$\curlyvee$$

$$\mu(\emptyset) = \dot{0}.$$

The union of disjoint sets of models is now both functional and associative. The results of these unions are nevertheless intuitively malformed. $\{v_1\}$, for example, is strictly less likely than both $\{v_2\}$ and $\{v_3\}$ yet the union of $\{v_1\}$ and $\{v_2\}$ is strictly larger than the union of $\{v_2\}$ and $\{v_3\}$.

The final non-logical restriction built into likelihood structures is that sentences or sets of models have a fixed magnitude regardless of which other inconsistent sentences or disjoint sets are combined with them.

M. *Monotonicity for Inconsistent Disjunction / Disjoint Union*
For any φ, ψ, γ where γ is inconsistent with both φ and ψ, $\mu(\varphi) \preceq \mu(\psi)$ if and only if $\mu((\varphi \vee \gamma)) \preceq \mu((\psi \vee \gamma))$.

Either adding or taking away a disjoint set does not change relative likelihood. Given a likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$, the examples above establish that

$$(\mathrm{N}) \wedge (\mathrm{C}) \wedge (\mathrm{I}) \wedge (\mathrm{F}) \not\Rightarrow (\mathrm{A}).$$

$$(\mathrm{N}) \wedge (\mathrm{C}) \wedge (\mathrm{I}) \wedge (\mathrm{F}) \wedge (\mathrm{A}) \not\Rightarrow (\mathrm{M}).$$

While instructive, the six restrictions presented in this section are not minimal. It is easy to prove, for example, that

$$(\mathrm{N}) \wedge (\mathrm{C}) \wedge (\mathrm{I}) \wedge (\mathrm{M}) \Rightarrow (\mathrm{F})$$

Accordingly, explicit commitment to (F) will be dropped in any context which also supposes (M).

While the set of restrictions (N), (C), (I), and (M) is minimal, whether or not

$$(\mathrm{N}) \wedge (\mathrm{C}) \wedge (\mathrm{I}) \wedge (\mathrm{M}) \Rightarrow (\mathrm{A})$$

is an open question. As the second example above shows, failures of (A) are consistent with a slight weakening of (M) identified by Friedman and Halpern (1995),

D. *Decomposition*
For any disjoint $A_1, B_1 \in \mathcal{F}_\mathcal{L}$ and disjoint $A_2, B_2 \in \mathcal{F}_\mathcal{L}$ such that $\mu(A_1) \leq \mu(A_2)$ and $\mu(B_1) \leq \mu(B_2)$, $\mu(A_1 \cup B_1) \leq \mu(A_2 \cup B_2)$.

At the same time, however, working through this example shows that the discrepancy between (M) and (D) is precisely what gives rise to the failure of associativity.

Proposition 2.4

Given a formal language L with Boolean connectives, a model space $\Omega_\mathcal{L}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$ satisfying (N), (C), (I), and (M), μ also satisfies

D. Decomposition
For any disjoint $A_1, B_1 \in \mathcal{F}_\mathcal{L}$ and disjoint $A_2, B_2 \in \mathcal{F}_\mathcal{L}$ such that $\mu(A_1) \leq \mu(A_2)$ and $\mu(B_1) \leq \mu(B_2)$, $\mu(A_1 \cup B_1) \leq \mu(A_2 \cup B_2)$.

F. Functionality for Inconsistent Disjunction / Disjoint Union
For any two pairs of inconsistent sentences φ, ψ and φ', ψ', if $\mu(\varphi) = \mu(\varphi')$ and $\mu(\psi) = \mu(\psi')$, then $\mu((\varphi \vee \psi)) = \mu((\varphi' \vee \psi'))$.

Proposition 2.5

For a formal language L with Boolean connectives, a model space $\Omega_\mathcal{L}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$,

(i) (N), (C), (I) $\not\Rightarrow$ (M).

(ii) (M), (C), (I) $\not\Rightarrow$ (N).

(iii) (N), (M), (I) $\not\Rightarrow$ (C).

(iv) (N), (C), (M) $\not\Rightarrow$ (I).

Given a likelihood assignment $\mu : \mathcal{F}_\mathcal{L} \to \mathbb{P}$ which satisfies the logical restrictions (N), (C), and (I) as well as the non-logical restrictions (A) and (M), $\langle \mathbb{P}|_{\mu[\mathcal{F}_\mathcal{L}]}, \circ \rangle$ where $\circ$ is the induced inconsistent disjunction operator is a likelihood structure.

Proposition 2.6

Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$ satisfying (N), (C), (I), (A), and (M), $\langle \mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}, \circ \rangle$ is a likelihood structure where $\circ : \mathbb{P} \times \mathbb{P} \twoheadrightarrow \mathbb{P}$ is defined by

$$\mu(\varphi) \circ \mu(\psi) = \mu((\varphi \vee \psi))$$

for $\varphi, \psi \in L$ inconsistent.

The definition of equivalence between likelihood spaces generalizes to a notion of equivalence across likelihood structures:

Definition 2.10 A likelihood structure $\langle \mathbb{P}_1, \circ_1 \rangle$ is *equivalent* to another likelihood structure $\langle \mathbb{P}_2, \circ_2 \rangle$ if and only if there exists a $\preceq$-preserving bijection $\delta : \mathbb{P}_1 \to \mathbb{P}_2$ such that for any $p, p' \in \mathbb{P}_1$, $\circ_2(\delta(p), \delta(p')) = \delta(\circ_1(p, p'))$.

Equivalently, we may also view each of $\langle \mathbb{P}_1, \circ_1 \rangle$ and $\langle \mathbb{P}_2, \circ_2 \rangle$ as models of the first-order language generated from $\langle \dot{0}, \dot{1}, \preceq, \circ \rangle$ in which case equivalence is again isomorphism between models in the typical sense.

The subjective Bayesian account, Jeffreys' (1967) infinitary proposal, a rational analogue of the subjective Bayesian account, and elimitivism about epistemic uncertainty all provide salient examples of likelihood structures:

$$\langle [0, 1], \leq, + \rangle$$

$$\langle [0, \infty], \leq, + \rangle$$

$$\langle \mathbb{Q} \cap [0, 1], \leq, + \rangle$$

$$\langle \{0, 1\}, \leq \rangle \text{ with } \circ (x, y) = \begin{cases} \uparrow & \text{if } x = 1 \text{ and } y = 1 \\ \max(x, y) & \text{otherwise} \end{cases}$$

Nothing, however, requires that a likelihood structure uses numerical objects; the interval approach entertained by Levi (1985) also generates a distinct likelihood structure:

$$\langle \{[a, b] \;:\; a, b \in [0, 1] \text{ with } a \leq b\}, [a_1, b_1] \leq [a_2, b_2] \Leftrightarrow a_1 \leq a_2 \text{ and } b_1 \leq b_2 \rangle$$
$$\text{with } \circ ([x_1, x_2], [y_1, y_2]) = \begin{cases} [x_1 + y_1, x_2 + y_2] & \text{if } x_2 + y_2 \leq 1 \\ \uparrow & \text{otherwise} \end{cases}$$

Taking all functions from $\mathcal{F}_{\mathcal{L}}$ to these likelihood structures which not only satisfy (N), (C), and (I) but also use the specified inconsistent disjunction operator provides a maximal collection of likelihood assignments that may be endorsed as rational credences. In the case of $\langle[0, 1], \leq, +\rangle$, this is at last the set of all finitely additive probability functions.

Accepting both the logical restrictions (N), (C), and (I) as well as the non-logical restrictions (A) and (M), theories of rational credence must provide a collection of functions from either L or $\mathcal{F}_{\mathcal{L}}$ to a likelihood structure $\langle\mathbb{P}, \circ\rangle$. Further, this collection must both refine the formal language's logic and combine likelihood values in an intuitive manner. Any theory of rational credence which falls short of this mark is either underspecified or incoherent according to whether it admits more than one such collection of functions or none.

Unfortunately, a particular collection of functions from a formal language L to a likelihood structure $\langle\mathbb{P}, \circ\rangle$ does not determine a corresponding theory of rational credence. It is entirely admissible for particular theories to posit that rational credence is dependent on some factor outside of L and $\Omega_{\mathcal{L}}$, e.g., a set of evidence or a prior probability function. Distinct theories may then endorse the same collection of functions from a particular language L to a particular likelihood structure $\langle\mathbb{P}, \circ\rangle$ but nevertheless disagree on the relationship between these further parameters and rational credence. In these circumstances, a formalization of these additional parameters is owed along with a collection of functions from L together with these parameters to a particular likelihood structure $\langle\mathbb{P}, \circ\rangle$. Note that such an account still determines a collection of functions from L to $\langle\mathbb{P}, \circ\rangle$ by allowing the additional parameters to vary freely. A well-defined collection of functions from L to a likelihood structure $\langle\mathbb{P}, \circ\rangle$ thus represents a necessary component of any theory of rational credence which accepts (A) and (M) even if it does not suffice to individuate all such theories.

In light of this, theories of rational credence may be usefully classified based on two criteria. First and foremost, by the collection of rational credences they endorse for a formal language L and model space $\Omega_{\mathcal{L}}$. Within this first criteria, it is useful to distinguish accounts which make use of only a single likelihood structure:

Definition 2.11 A theory of rational credence is *categorical* if and only if it fixes a single likelihood structure $\langle\mathbb{P}, \leq, \circ\rangle$ for any formal language L and model space $\Omega_{\mathcal{L}}$.

Contemporary accounts of rational credence are not only almost exclusively categorical but also in widespread agreement on the use of the $\langle[0, 1], \leq, +\rangle$ likelihood structure. As the current section shows, neither of these is delivered by purely logical considerations nor even by the intuitive restrictions encoded in (A) and (M). Second, theories of rational credence may be usefully classified by whether or not they endorse any further parameters, and if so, the exact structure and nature of these parameters. Even if we are in agreement that $\langle[0, 1], \leq, +\rangle$ is the correct likelihood structure and that all probability functions are possible for propositional languages, there may remain significant disagreement about whether or not a particular set of evidence from a propositional language determines a unique probability function as the rational credence. Contemporary debates around rational credence tend to fall into this second category; the next chapter argues that there is good reason to be worried over the first.

2.3 Connecting Assignments and Comparative Relations

At the outset of this chapter, two formalisms for ranking objects were presented: absolute and comparative rankings. We come full circle in this section and connect the previous section's work on likelihood spaces and structures with better known work on comparative likelihood rankings.

2.3.1 NCI-Assignments and C01a23bN-Relations

Assignments into a likelihood space satisfying each of *Normativity of Logic* (N), *Certain Validities* (C), and *Impossible Contradictions* (I) correspond to comparative likelihood relations satisfying (C0), (C1a), (C2), (C3b), and (CN)—or, more briefly, C01a23bN-relations.

Proposition 2.7

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, $\mathbb{P}$ a likelihood space, $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ an NCI-likelihood assignment. Then, the relation $\precsim_\mu$ over $\mathcal{F}_{\mathcal{L}}$ defined by

$$A \precsim_\mu B \Leftrightarrow \mu(A) \leq \mu(B)$$

satisfies

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

CN. Normativity of Logic.
If $A \subseteq B$, then $A \precsim B$.

Definition 2.12 Given a formal language L and a model space $\Omega_{\mathcal{L}}$ for L, a likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ *represents*[7] a binary relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ just in case for any $A, B \in \mathcal{F}_{\mathcal{L}}$,

$$A \precsim B \text{ if and only if } \mu(A) \leq \mu(B).$$

Definition 2.13 Given a formal language L and a model space $\Omega_{\mathcal{L}}$ for L, a binary relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ is *representable* in a likelihood space $\mathbb{P}$ just in case there exists a likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ such that for any $A, B \in \mathcal{F}_{\mathcal{L}}$,

$$A \precsim B \text{ if and only if } \mu(A) \leq \mu(B).$$

Proposition 2.8

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

CN. Normativity of Logic
If $A \subseteq B$, then $A \precsim B$.

Then, there exists a likelihood space $\mathbb{P}_{\precsim}$ ("the likelihood space induced by $\precsim$") such that $\precsim$ is representable in $\mathbb{P}_{\precsim}$ by an NCI-Likelihood Assignment.

Definition 2.14 Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a C01a23bN-relation $\precsim$ defined over $\mathcal{F}_{\mathcal{L}}$, *the likelihood space induced by* $\precsim$ is

- domain $\mathcal{F}_{\mathcal{L}}/\sim$
- $\dot{0}^{\mathbb{P}_{\precsim}} = [\emptyset]$
- $\dot{1}^{\mathbb{P}_{\precsim}} = [\Omega_{\mathcal{L}}]$
- $\leq^{\mathbb{P}_{\precsim}} = \{\langle [A], [B] \rangle \mid A \precsim B\}$.

[7] This notion is 'agrees' in Fine (1973) and 'strictly agrees' in Savage (1972).

Definition 2.15 A likelihood space $\mathbb{P}$ embeds in a likelihood space $\mathbb{P}'$ if and only if there exists an injection $\delta : \text{dom}(\mathbb{P}) \to \text{dom}(\mathbb{P}')$ such that

$$\delta(\dot{0}^{\mathbb{P}}) = \dot{0}^{\mathbb{P}'},$$

$$\delta(\dot{1}^{\mathbb{P}}) = \dot{1}^{\mathbb{P}'},$$

and for any $p_1, p_2 \in \mathbb{P}$,

$$p_1 \precsim^{\mathbb{P}} p_2 \text{ if and only if } \delta(p_1) \precsim^{\mathbb{P}'} \delta(p_2).$$

Definition 2.16 A likelihood structure $\langle \mathbb{P}, \circ \rangle$ embeds in a likelihood structure $\langle \mathbb{P}', \circ' \rangle$ if and only if there exists an injection $\delta : \text{dom}(\mathbb{P}) \to \text{dom}(\mathbb{P}')$ such that

$$\delta(\dot{0}^{\mathbb{P}}) = \dot{0}^{\mathbb{P}'},$$

$$\delta(\dot{1}^{\mathbb{P}}) = \dot{1}^{\mathbb{P}'},$$

and for any $p_1, p_2 \in \mathbb{P}$,

$$p_1 \precsim^{\mathbb{P}} p_2 \text{ if and only if } \delta(p_1) \precsim^{\mathbb{P}'} \delta(p_2),$$

$$\delta(p_1 \circ^{\mathbb{P}} p_2) = \delta(p_1) \circ^{\mathbb{P}'} \delta(p_2).$$

Proposition 2.9

> *Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$. If $\precsim$ is representable in a likelihood space $\mathbb{P}$ and $\mathbb{P}$ embeds into another likelihood space $\mathbb{P}^+$, then $\precsim$ is representable in $\mathbb{P}^+$. Further, if $\precsim$ is representable in a likelihood structure $\langle \mathbb{P}, \circ \rangle$ and $\langle \mathbb{P}, \circ \rangle$ embeds into another likelihood structure $\langle \mathbb{P}^+, \circ^+ \rangle$, then $\precsim$ is representable in $\langle \mathbb{P}^+, \circ^+ \rangle$.*

Definition 2.17 Let a formal language L with Boolean connectives and model space $\Omega_{\mathcal{L}}$ be given. A likelihood space $\mathbb{P}$ is *the minimal representing space for a relation $\precsim$ over $\mathcal{F}_{\mathcal{L}}$* if and only if $\precsim$ is representable in $\mathbb{P}$ and for any likelihood space $\mathbb{P}'$ capable of representing $\precsim$, there exists an embedding of $\mathbb{P}$ into $\mathbb{P}'$.

Proposition 2.10

Let L be a formal language, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

CN. Normativity of Logic
If $A \subseteq B$, then $A \precsim B$.

Then, $\precsim$ is representable in a likelihood space $\mathbb{P}$ by an NCI-likelihood assignment if and only if $\mathbb{P}_{\precsim}$ embeds into $\mathbb{P}$.

2.3.2 NCIAM-Assignments and C01a23b4A-Relations

Assignments into a likelihood space satisfying each of *Normativity of Logic* (N), *Certain Validities* (C), *Impossible Contradictions* (I), *Associativity* (A), and *Monotonicity* (M) correspond to comparative likelihood relations satisfying C01a23b4A.

Proposition 2.11

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, $\mathbb{P}$ a likelihood space, $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ an NCIAM-likelihood assignment. Then, the relation $\lesssim_\mu$ over $\mathcal{F}_{\mathcal{L}}$ defined by

$$A \lesssim_\mu B \Leftrightarrow \mu(A) \leq \mu(B)$$

satisfies

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \lesssim A$.

C2. Transitivity
If $A \lesssim B$ and $B \lesssim C$, then $A \lesssim C$.

C3b. Boundedness
$\bot \lesssim A \lesssim \Omega_{\mathcal{L}}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \lesssim B$ if and only if $A \cup C \lesssim B \cup C$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, and $G_4 \sim F_3 \cup E_3$ with matching subscripts disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Despite its complexity, the comparative associativity condition above is a direct translation of associativity for $\circ$ into the comparative setting.

$$\begin{array}{ccccccccccccccc}
(& p_1 & \circ & p_2 &) & \circ & p_3 & = & p_1 & \circ & (& p_2 & \circ & p_3 &) \\
 & & C_2 & & & & D_2 & & H_4 & & & & G_4 & & \\
 & A_1 & & B_1 & & & & & & & & E_3 & & F_3 &
\end{array}$$

C_2 is thus a collection with likelihood $p_1 \circ p_2$ while A_1 is a collection with likelihood p_1 and B_1 is a collection with likelihood p_2. Note that sets with matching subscripts are required to be disjoint in the comparative formulation so that the $\circ$ operation may be applied.

Definition 2.18 Given a formal language L and a model space $\Omega_{\mathcal{L}}$ for L, a binary relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ is *representable* in a likelihood structure $\langle \mathbb{P}, \circ \rangle$ just in case there exists a likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ such that for any $A, B \in \mathcal{F}_{\mathcal{L}}$,

$$A \precsim B \text{ if and only if } \mu(A) \leq \mu(B)$$

and for any disjoint sets $C, D \in \mathcal{F}_{\mathcal{L}}$,

$$\mu(C \cup D) = \mu(C) \circ \mu(D).$$

Proposition 2.12

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, $G_4 \sim F_3 \cup E_3$, and all sets with matching subscripts are disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Then, there exists a likelihood structure $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ ("the likelihood structure induced by $\precsim$") such that $\precsim$ is representable in $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$.

Definition 2.19 Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a C01a23b4A-relation $\precsim$ defined over $\mathcal{F}_{\mathcal{L}}$, *the likelihood structure* $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ *induced by* $\precsim$ is

- domain $\mathcal{F}_{\mathcal{L}} / \sim$
- $\dot{0}^{\mathbb{P}_{\precsim}} = [\emptyset]$
- $\dot{1}^{\mathbb{P}_{\precsim}} = [\Omega_{\mathcal{L}}]$
- $\preceq^{\mathbb{P}_{\precsim}} = \{\langle [A], [B] \rangle \mid A \precsim B\}$
- $[A] \circ_{\precsim} [B] \downarrow= [C]$ if and only if there exists disjoint $A', B' \in \mathcal{F}_{\mathcal{L}}$ such that $A' \in [A], B' \in [B]$, and $A' \cup B' \in [C]$.

Definition 2.20 Let a formal language L with Boolean connectives and model space $\Omega_{\mathcal{L}}$ be given. A likelihood structure $\langle \mathbb{P}, \circ \rangle$ is *the minimal representing structure for a relation* $\precsim$ *over* $\mathcal{F}_{\mathcal{L}}$ if and only if $\precsim$ is representable in $\langle \mathbb{P}, \circ \rangle$ and for any likelihood structure $\langle \mathbb{P}', \circ' \rangle$ capable of representing $\precsim$, there exists an embedding of $\langle \mathbb{P}, \circ \rangle$ into $\langle \mathbb{P}', \circ' \rangle$.

Proposition 2.13

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, $G_4 \sim F_3 \cup E_3$, and all sets with matching subscripts are disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Then, $\precsim$ is representable in $\langle \mathbb{P}', \circ' \rangle$ if and only if $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ embeds into $\langle \mathbb{P}', \circ' \rangle$.

Chapter 3

The Problem with Probabilism

Since the credences of actual agents are largely consistent in only their lack of consistency, rational credence is the natural focus of theoretical discussion. The mainstream conception of rational credence is and always has been dominated by probabilism:

- *Probabilism*
  If c is a rational credence over the event space $\mathcal{F}$, then c is[1] a probability function.

Note that probabilism only provides a necessary condition on rational credences; it is perfectly consistent to also maintain, for example, that there is a unique rational credence in a given circumstance. Probabilism is thus a strictly weaker commitment than either subjective or objective Bayesianism. A large number of arguments have been offered in support of probablism, most famously (i) the Dutch book argument, (ii) representation arguments from comparative probability, and (iii) gradational accuracy arguments. The current chapter argues that none of these provides an adequate justification.

The root of this inadequacy is a systematic failure to justify the use of real numbers. This is most immediately apparent with the following two consequences of probabilism:

- *Comparability for Rational Credences*
  For any rational credence c and propositions φ, ψ, either $c(\varphi) \leq c(\psi)$ or $c(\psi) \leq c(\varphi)$.

- *Archimedean Property for Rational Credences*
  For any rational credence c and proposition φ, either $c(\varphi) = \dot{0}$ or there exists $n \in \mathbb{N}$ such that $c(\varphi)^n = \underbrace{c(\varphi) \circ c(\varphi) \circ \cdots \circ c(\varphi)}_{n \text{ times}}$ is undefined.

Neither comparability nor the Archimedean property appear as explicit axioms in the usual presentation of probability theory:

[1] I follow Pettigrew (2016) and Bradley (2017) here. It is not unusual, however, to see 'is represented by', 'can be represented' (Caie 2013), or even 'can be faithfully represented by' (Easwaran and Fitelson 2012). None of these reformulations escape the shortcomings noted in this chapter.

Definition 3.1 A *probability space* is a triple $\langle \Omega, \mathcal{F}, \mu \rangle$ where Ω is a non-empty set ("the state space"), $\mathcal{F}$ is a σ-algebra[2] ("the event space"), and μ is a function ("the probability function") from $\mathcal{F}$ to $\mathbb{R}$ satisfying the following three axioms:

K1. *Normality*
$\mu(\Omega) = 1$;

K2. *Non-Negativity*
For any $A \in \mathcal{F}$, $\mu(A) \geq 0$;

K3. *Countable Additivity*
For a countable collection $\{A_i\}$ of pairwise disjoint sets from $\mathcal{F}$,

$$\mu(\cup\{A_i\}) = \sum_i \mu(A_i).$$

Indeed, neither property is entailed by the combination of (K1), (K2), and (K3).

Probabilism is nevertheless committed to both comparability and the Archimedean property for rational credences by the stipulation that μ maps into the real numbers. In the case of comparability, this is obvious. Since the real numbers themselves satisfy comparability, using any subset of the real numbers as our likelihood space immediately imposes the same structure on rational credences. The situation with the Archimedean property is essentially similar. If our likelihood space is stipulated to be any subset of the real numbers and the disjunction of inconsistent sentences is identified with addition—i.e., a likelihood structure $\langle P, + \rangle$ for $P \subseteq \mathbb{R}$ is adopted—then rational credences immediately inherit the general features of this structure. Since for any $r \in \mathbb{R}$, $n \times r > \dot{1}$ for some $n \in \mathbb{N}$,[3] among these general features is the Archimedean property, viz. a prohibition against infinitesimal values. The stipulation that probability functions map into the real numbers thus introduces additional commitments. Probabilism is committed to substantially more than just (K1), (K2), and (K3).

While widespread,[4] explicit focus on (K1), (K2), and (K3) alone as 'the axioms of probability theory' is thus deeply misleading. A complete axiomatization of probability theory includes an additional entry for the initial restriction on each component of a probability space:

KΩ. Ω is a non-empty set.

K$\mathcal{F}$. $\mathcal{F}$ is a σ-Algebra over Ω.

---

[2] A σ-algebra over a set Ω is a subset of $\mathcal{P}(\Omega)$ containing $\emptyset$ which is closed under complement, countable union, and countable intersection.

[3] Note that the certain value $\dot{1}$ need only be some real number r^* here.

[4] See, for example, the presentation of probabilism in Bradley (2017), Pettigrew (2016), Caie (2013), Joyce (1998, 2009), and Jeffrey (1986).

Kμ. μ is a function from $\mathcal{F}$ to $\mathbb{R}$.

The first of these additional constraints is, by all accounts, unproblematic. While the infinitary nature of (K$\mathcal{F}$) has attracted sporadic criticism in combination with (K3), the difficulties here pale in comparison to those introduced by (Kμ). $\mathbb{R}$ (or the intended model of the real numbers), while familiar, is not uniquely identified by any first-order theory. Since (Kμ) explicitly invokes $\mathbb{R}$, this complexity is inherited by the probability formalism. Probabilism thus embraces a second-order axiom in the form of (Kμ) while only drawing explicit attention to (K1), (K2), and (K3).

Intentional or not, contemporary arguments for probabilism have perpetuated and played on this sleight of hand to the point of being actively misleading. The Dutch book argument, representation arguments from comparative probability, decision-theoretic representation arguments, and gradational accuracy arguments all systematically fail to justify comparability and the Archimedean property much less (Kμ) as a whole. In isolation, this shortcoming exposes a lacuna in contemporary accounts and calls into question the consensus on probabilism. Criticisms of comparability and the Archimedean property are not, however, novel. Influential commentaries like Keynes (1921) and Koopman (1940) explicitly reject comparability while equally important entries like Koopman (1940), Shimony (1955), Kemeny (1955), and Lewis (1981) reject the Archimedean property under one guise or another. The situation only worsens in the current literature with a large number of advocates for incomparability in the form of imprecise credences (Joyce 2010) as well as proponents of explicitly non-Archimedean conceptions of probability (Narens 2007; Benci et al. 2013). This long history suggests that defenders of probabilism have not so much overlooked this weakness in their account as deliberately ignored it. I conclude that the case for probabilism has been systematically overstated.

3.1 The Dutch Book Argument

The Dutch book argument first explicitly appears in Ramsey (1931): "If anyone's mental condition violated these laws [of probability], his choice would depend on the precise form in which the options were offered him, which would be absurd. He could have a book made against him by a cunning better and would then stand to lose in any event" (182). After Ramsey's premature death, this argument was independently redeveloped and then popularized by de Finetti ([1937] 1980) with substantive refinements from Kemeny (1955), Lehman (1955), and Shimony (1955). In what will quickly become a running theme, the most egregious shortcomings of the Dutch book argument can be readily identified by focusing on the means by which comparability and the Archimedean property are secured.

3.1.1 Terminology and Argument

Both the Dutch book argument and the surrounding discussion are steeped in gambling terminology. A *bet on an event E for a stake S* is an ordered pair $\langle S - qS, -qS \rangle$ where

$S - qs$ is paid out in the case E is true and $-qS$ otherwise. Diagrammatically,

	E	$\neg E$
Bet on E for stake S	$S - qS$	$-qS$

The product qS is known as the *betting price* while q in isolation is the *betting quotient* or betting ratio. Note that, for a particular stake S, betting prices can be easily derived from betting quotients and betting quotients can be easily derived from betting prices. Finally, both the terminology and payouts given above are particularly natural if one imagines paying qS for a bet which returns S if E and 0 otherwise.

Given this framework, it's natural to talk not only about bets $\langle S - qS, -qS \rangle$ on E but also about the corresponding bet against E, $\langle -S + qS, qS \rangle$. Bets on and bets against simply reverse gain and loss relative to a fixed stake S:

	E	$\neg E$
Bet on E	$S - qS$	$-qS$
Bet against E	$-S + qS$	qS

Where bets on E are naturally thought of as buying a bet for qS which pays S if E and nothing otherwise, bets against E correspond to selling a bet for qS which pays S if E and nothing otherwise. For a fixed betting quotient q, note that a bet against E is the same as a bet on E for stake $-S$ but distinct from a bet on $\neg E$ for stake S.

Finally, for a collection of betting quotients, a *Dutch book* is a combination of bets (either for or against events) which ensures a net positive outcome. Such a combination requires that the inverted combination (replacing bets for with bets against and vice-versa) is a sure loss. Given the structure of the Dutch book argument, commentators often adopt the perspective of the side suffering a sure-loss and talk of a Dutch book being made against an agent or being 'Dutch bookable'. While it plays no role in the following discussion, it is common—following de Finetti ([1937] 1980)—to call a collection of betting quotients *coherent* if and only if there does not exist a Dutch book at those betting quotients.

The Dutch book argument ostensibly forges a link between probabilistic credences and immunity to Dutch books. Let a σ-algebra $\mathcal{F}$ over a nonempty set Ω be given. The contemporary Dutch book argument can be summarized as follows:

The Dutch Book Argument:

(P1) If an agent has credence c over $\mathcal{F}$, then they endorse $q = c$ as betting quotients over $\mathcal{F}$.

(P2) For any collection of betting quotients q over $\mathcal{F}$, there exists a Dutch book with respect to q if and only if q is not a probability function.

(P3) If there exists a Dutch book under a collection of betting quotients q, then it is irrational to endorse these betting quotients.

(C) If c is a rational credence over an algebra $\mathcal{F}$, then c is a probability function.

Each of the premises here bears some comment. The first premise (P1) postulates an indirect connection between credences and betting. Agents may, of course, choose to bet in a manner which does not conform to their credences for any number of reasons. Nevertheless, there is a natural sense in which a credence c endorses the particular betting quotients $q = c$, viz. supposing utility linear in dollars and evaluation of bets by expectation, these are the betting quotients for which betting prices accord with the agent's own valuation.[5] Following Christensen (1996), this—rather than any explicit act on the part of the agent—is the sense of the endorsement in (P1).

The second premise (P2) is intended as a purely mathematical result and sometimes separated into two components:

- *The Dutch Book Theorem:* For any collection of betting quotients q over $\mathcal{F}$, there exists a Dutch book with respect to q if q is not a probability function.
- *The Converse Dutch Book Theorem:* For any collection of betting quotients q over $\mathcal{F}$, there exists a Dutch book with respect to q only if q is not a probability function.

While the validity of the Dutch book argument only requires the former, the converse result blocks possibilities which might call the truth of (P1) and (P3) into question, most obviously if no betting quotients were immune to Dutch book (Hájek 2009; Hájek 2008). Since the converse result is relevant to the overarching discussion and there is no loss in opting for a stronger mathematical result, I make use of the combined result.

Issues with the third premise (P3) of the Dutch book argument almost exclusively revolve around the meaning of 'irrational'. In particular, it is tempting to read 'irrational' here as something like 'ill-advised on pragmatic grounds'. This reading, however, is doubly problematic. First, it may sometimes be pragmatically rational to endorse betting quotients susceptible to Dutch book if one has an overriding incentive or the other bettors are not intelligent enough to take advantage. Supposing the pragmatic reading, (P3) is then false

[5] These are the *fair* odds in the terminology of Christensen (1996).

as stated. Second, the pragmatic reading appears to sever any connection to epistemology; Kennedy and Chihara (1979) provide a classic statement of the issue:

> The factors that are supposed to make it irrational to have an inadmissible set of beliefs [in these betting situations] are entirely irrelevant, epistemologically, to the truth of the propositions in question. The fact (if it is a fact) that one will be bound to lose money unless one's degrees of belief are admissible just isn't epistemologically relevant to the truth of those beliefs. (30)

Following Skyrms (1980) and Christensen (1996), both problems can be avoided with a non-pragmatic reading of (P3). In particular, we ought to maintain that the existence of a Dutch book exhibits an inconsistency of judgment which is incompatible with endorsement by an epistemically rational agent.

Put together, (P1), (P2), and (P3) appear to require that any rational agent's credence is a probability function. Not because such agents are required to bet at their credences or because they might lose money if they do not, but because credences endorse particular betting quotients in a simple sort of betting situation and these betting quotients are inconsistent if they are not probabilistic. The sin here is not the possibility of a sure loss; it is the inconsistent valuations which lead to the sure loss.

3.1.2 The Problem with the Dutch Book Argument

The Dutch book argument is not sound. The culprit, however, is neither (P1) nor (P3); it is the 'mathematical theorem' (P2). There is a theorem which looks very much like (P2):

Theorem 3.1 Kemeny (1955); Lehman (1955)

For any σ-algebra $\mathcal{F}$ and any collection of betting quotients $q : \mathcal{F} \to \mathbb{R}$, there exists a Dutch book with respect to q if and only if q is not a probability function.

Unlike (P2), however, this theorem assumes that betting quotients are real numbers; q is a function *to* $\mathbb{R}$ and thus already satisfies both comparability and the Archimedean property. Without this guarantee, probability functions are no longer the only likelihood assignments which are immune to Dutch books, and so (P2) is false.

A minimal counterexample can be obtained by setting $\Omega = \{w_1, w_2\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, and betting quotients

$$\begin{aligned} q(\emptyset) &= 0 \\ q(\{w_1\}) &= \frac{1}{2}_a \\ q(\{w_2\}) &= \frac{1}{2}_b \\ q(\{w_1, w_2\}) &= 1 \end{aligned}$$

where $\frac{1}{2}_a$ and $\frac{1}{2}_b$ are incomparable to one another but otherwise behave as $\frac{1}{2}$. Bets for or against either Ω or $\emptyset$ then behave as usual. The events $\{w_1\}$ and $\{w_2\}$, however, are somewhat odd in that there is no price which produces indifference between bets for and bets against. Any price strictly below $\frac{1}{2}S$ is acceptable with a bet for either $\{w_1\}$ or $\{w_2\}$ while any price strictly above $\frac{1}{2}S$ is acceptable for a bet against. Bets costing exactly $\frac{1}{2}S$ for or against either event, however, are uniformly rejected. Since any sequence of acceptable bets with these quotients is also acceptable to an agent who sets $q(\{w_1\}) = q(\{w_2\}) = \frac{1}{2}$, these betting quotients are both non-probabilistic and immune to Dutch books. It follows immediately that (P2) is false.

This discrepancy between the actual Dutch book theorem and the premise invoked in the Dutch book argument is widespread. Compare, for instance, the premise provided by Baillie (1973, 393):

> "[betting quotients] are coherent only if a probability."

or Kennedy and Chihara (1979, 20-21):

> "a set of betting quotients is coherent if and only if it [conforms to the probability calculus]."

or Gillies (2000, 59):

> "A set of betting quotients is coherent if and only if they satisfy the axioms of probability."

The Dutch book theorem guarantees only that *real-valued* betting quotients which avoid Dutch books are probabilistic. The example above proves that this result does not generalize to all betting quotients, and so each premise is false as stated.

Commentators explicitly concerned with the Dutch book theorem tend to avoid this mistake. Each of Lehman (1955), Kemeny (1955), and Shimony (1955) for example explicitly recognize a restriction to real numbers. Surveying these presentations suggests two strategies for repairing (P2); neither strategy, however, allows for the Dutch book argument to function as a justification for probabilism. First, we might follow Lehman (1955) and Shimony (1955) in simply stipulating that rational credences are real-valued. In light of (P1), this makes replacing (P2) with the Dutch book theorem proper unproblematic. Unfortunately, whether or not rational credences are real-valued is precisely the point at issue. This strategy is clearly question begging in a justification for probabilism ex nihilo.

Second, we may follow Kemeny (1955) and derive that betting quotients are real-valued by supposing a single betting price over both bets on and bets against an event E.[6] Since betting prices are sums of money, the resulting betting quotients must then be real-valued.[7]

[6] Kemeny (1955) is not quite so straightforward. The crucial combination in the text is fixing a betting quotient for every event E (263) and supposing that the resulting qS and $(1-q)S$ are actual sums of money (265). For any S and E, this entails a single betting price over both bets on and bets against E.

[7] Indeed, betting quotients are rational-valued if the monetary system admits of a smallest unit.

Kemeny's requirement thus limits the possible structure of betting quotients by requiring expressibility as a ratio of two monetary values. Even if we are willing to accept this sort of constraint on betting quotients, it merely pushes the current difficulties onto (P1). If betting quotients are necessarily real-valued, credences can only be endorsed as betting quotients if they too satisfy both comparability and the Archimedean property. Any failure of comparability or the Archimedean property in our credences then immediately entails that (P1) is false since no corresponding collection of betting quotients exists to be endorsed. As a result, this strategy too fails to provide us with a coherent argument for probabilism ex nihilo.

Despite its fame, the Dutch book argument falls short of its intended conclusion. This failure is not, however, a result of either (P1) or (P3), the traditional points of criticism. The flaw in the argument is instead to be found with (P2). The restriction to real-valued betting quotients in the Dutch book theorem proper is critical for the result. By dropping this restriction in (P2), the Dutch book argument exchanges a mathematical theorem for a straightforward falsehood. Unfortunately, the validity of the Dutch book argument requires this stronger claim. Replacing (P2) with the Dutch book theorem proper, we may at best obtain a conditional conclusion: if rational credences are real-valued, then they must be probability functions.

3.2 Representation Arguments

Outside of the Dutch book argument, the most ubiquitous arguments for probabilism are representation arguments. Ideally, representation arguments for probabilism advance in two stages. First, a set of intuitively true rationality axioms are identified in some domain. Second, a formal theorem is proved which shows that satisfying these rationality axioms is necessary and sufficient for probabilism. The intuitively true rationality axioms demand as it were that rational credences are probability functions. This section presents the most straightforward attempt to implement this strategy. Despite their fame, actual representation arguments fall far short of the ideal.

3.2.1 Representation Argument from Finite Comparative Likelihood

The most straightforward representation arguments target comparative likelihood. Where probability functions provide an absolute, quantitative measure of likelihood, the comparative likelihood relation makes only binary comparisons:

$$\varphi \precsim \psi \text{ if and only if } \psi \text{ is at least as likely as } \varphi.$$

If an intuitive set of rationality axioms for the comparative likelihood relation $\precsim$ can be shown to guarantee equivalence to a probability function, then we will have an instance of the ideal representation argument and a powerful justification for probabilism.

For a nonempty set of possibilities Ω, an algebra $\mathcal{F}$ over Ω, and $A, B, C \in \mathcal{F}$, a number of simple restrictions on $\precsim$ are clearly necessary for a correspondence with probability functions:

C0. *Nontriviality*
$\emptyset \prec \Omega$.

C1a. *Reflexivity*
$A \precsim A$.

C1b. *Comparability; Totality; Connected*
$A \precsim B$ or $B \precsim A$.

C2. *Transitivity*
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3a. *Nonnegativity*
$\perp \precsim A$.

C3b. *Boundedness*
$\perp \precsim A \precsim \Omega$.

(C1a) and (C2) entail that $\precsim$ is a preorder while (C1b) and (C2) entail that $\precsim$ is a total preorder.

Definition 3.2 A *preorder* over a space $\mathcal{F}$ is a reflexive and transitive binary relation; a *total preorder* over a space $\mathcal{F}$ is a preorder which also satisfies comparability.

Preorders are so called because of their close relationship with orderings; given a preorder $\precsim$ over a space $\mathcal{F}$, an ordering $\leq$ is induced over $\mathcal{F}/\sim$ by taking $[a] \leq [b]$ if and only if $a \precsim b$. Adding (C0) and (C3) as additional restrictions on $\precsim$ produces a total preorder with $\emptyset$ as a minimal element and Ω as a maximal element.

It is nevertheless easy to construct C01a1b23ab-preorders which are both implausible for a rational agent and non-probabilistic:

$$\emptyset \prec A \cup B \prec A \prec B \prec \Omega.$$

It is clearly illicit for the union operation to reduce comparative likelihood as in the example above. In response, de Finetti ([1937] 1980) proposes the additional requirement that unions with a disjoint set preserve inequalities:

C4. *Monotonicity; Additivity*
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

De Finetti ([1937] 1980) famously conjectured that any C01b23b4 comparative likelihood relation over an algebra $\mathcal{F}$ on a finite set of possibilities Ω is representable by a probability function; that these rationality axioms were all that was needed in the finite case. An example due to Kraft et al. (1959) proves that this is not the case:

$$
\begin{aligned}
\emptyset &\prec a \prec b \prec c \prec ab \prec ac \prec d \prec ad \prec bc \prec e \prec abc \prec bd \prec cd \prec ae \prec abd \\
&\prec be \prec acd \prec ce \prec bcd \prec abe \prec ace \prec de \prec abcd \prec ade \prec bce \prec abce \\
&\prec bde \prec cde \prec abde \prec acde \prec bcde \prec abcde = \Omega.
\end{aligned}
$$

C01b23b4 comparative likelihood relations need not fix a consistent size for all possibilities.

To bridge the gap between C01b23b4 likelihood relations and probability functions, Scott (1964) introduces

C5$_\mathbf{S}$. *Finite Cancellation*
If two sequences of sets from $\mathcal{F}$, $\langle A_1, \ldots, A_N \rangle$ and $\langle B_1, \ldots, B_N \rangle$, contain each possibility in Ω the same number of times and $A_i \precsim B_i$ for all $i < N$, then $B_N \precsim A_N$.

For a finite set of possibilities Ω, the collection C01b3a5$_S$ is sufficient to guarantee representation by a probability function.

Theorem 3.2 Scott (1964)

For any algebra $\mathcal{F}$ over a finite set Ω and any binary relation $\precsim$ on $\mathcal{F}$, $\precsim$ is representable by a probability function $\mu : \mathcal{F} \to [0, 1]$ if and only if $\precsim$ satisfies C01b3a5$_S$.

Note that all of the remaining axioms introduced in this section—(C1a), (C2), (C3b), and (C4)—follow from C01b3a5$_S$ even without the restriction to finite Ω.

Proposition 3.1

Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation over $\mathcal{F}$. Then,

(i) (C1b) $\Rightarrow$ (C1a)

(ii) (C5$_S$) $\Rightarrow$ (C2)

(iii) (C3a), (C5$_S$) $\Rightarrow$ (C3b)

(iv) (C5$_S$) $\Rightarrow$ (C4).

Comparative likelihood relations satisfying C01b3a5$_S$ are comparative likelihood relations satisfying C01ab3ab45$_S$.

Though seldom discussed, an initial representation argument for finite cases is now available:

C01B3A5$_S$ REPRESENTATION ARGUMENT

(P1) Each of (C0), (C1b), (C3a), and (C5$_S$) are rational restrictions on comparative likelihood relations.

(P2) For any finite Ω, a comparative likelihood relation satisfies C01b3a5$_S$ if and only if it is representable by a probability function $\mu : \mathcal{F} \to [0, 1]$.

(C) If c is a rational credence over an algebra $\mathcal{F}$ with finite Ω, then c is a probability function.

The first premise (P1) is intended to be intuitively true while the second premise (P2) is Scott (1964)'s theorem. Put together, we have that rational credences are probability functions whenever Ω is finite.

3.2.2 The Problem with the Argument from Finite Comparative Likelihood

There are two clear problems with the C01b3a5$_S$ representation argument. First and most egregiously, the argument has already begged the question in favor of probabilism. Comparability was one of two dubious properties singled out at the start of this chapter, and yet it appears in (P1) as a requirement of rationality. The comparison with (C0), (C1a), (C3a), or (C3b) here is telling. Per chapter 2, each of these restrictions follow from the assumption that rational credences extend logic; any NCI-likelihood assignment gives rise to a comparative relation satisfying (C0), (C1a), (C3a), and (C3b). The correspondence results proved there establish that neither the comparability axiom (C1b) nor Scott (1964)'s (C5$_S$) rise to this same standard. Instead, the primary motivation for declaring comparability a requirement of rationality appears to be all and only that this is necessary in order to establish a correspondence with probability functions. We thus have good reason to reject (P1).

Second, the conclusion of the C01b3a5$_S$ representation argument does not follow from the premises; we are entitled only to the conclusion that c is representable by a probability function. In general, representability in a likelihood space $\mathbb{P}$ or likelihood structure $\langle \mathbb{P}, \circ \rangle$ does not imply non-representability in other spaces or structures. This is obviously true for particular instances. For any particular choice of Ω, the comparative likelihood relation

$$\emptyset \prec \Omega$$

is, for example, representable in every likelihood space and likelihood structure. More surprisingly, establishing necessary and sufficient conditions for representability in some likelihood space $\mathbb{P}$ or likelihood structure $\langle \mathbb{P}, \circ \rangle$ also does not guarantee that this relationship is unique. Notational variants—e.g., the structure $\langle [1,2], 1, 2, \leq, \circ(x,y) = x+y-1 \rangle$ in lieu of the conventional probability structure—are a trivial counterexample. If these were the only difficulties, then the stronger conclusion—rational credences are probability functions—would be defensible. This phenomenon also extends, however, to non-isomorphic likelihood spaces and structures:

Proposition 3.2

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$. $\precsim$ is representable by an NCI-likelihood assignment into $\langle \mathbb{Q} \cap [0,1], 0, 1, \leq, + \rangle$ if and only if $\precsim$ satisfies C01b3a5$_S$.

Theorem 3.3 Narens (2007)

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a finite model space for L, $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$, and $^*[0,1]$ a non-standard extension of the unit interval. $\precsim$ is representable by an NCI-likelihood assignment into $\langle {}^*[0,1], 0, 1, \leq, + \rangle$ if and only if $\precsim$ satisfies C01b3a5$_S$.

The formal result underpinning (P2) is thus not nearly so impressive as it first appears. The "rationality" axioms C01b3a5$_S$ are consistent with both the truth and falsity of probabilism over finite sets.

The multiplicity of representation theorems here stems from the restriction to finite sets. Representation theorems establish only that the relevant likelihood assignments do not go awry over the specified domain. Disagreements between likelihood spaces or structures which fall outside this domain (a countably infinite number of values versus continuum many, the existence or non-existence of infinitesimal values) fail to manifest, and so we have a proliferation of representations. The relevant interactions here can be brought to the fore with an extreme example. It is not difficult to establish a representation theorem for C01b3a5$_S$ likelihood relations and the likelihood structure which denies any intermediate values:

Proposition 3.3

Let L be a formal language, $\Omega_{\mathcal{L}}$ a model space for L which contains only a single model, and $\precsim$ a binary relation over $\mathcal{F}_{\mathcal{L}}$. Then, $\precsim$ satisfies C01b3a5$_S$ if and only if it is representable in

$$\langle \{0,1\}, 0, 1, \leq \rangle \textit{ with } \circ(x,y) = \begin{cases} \uparrow & \textit{if } x = 1 \textit{ and } y = 1 \\ \max(x,y) & \textit{otherwise} \end{cases}.$$

The trick is simply to eliminate all the problematic cases using restrictions on L or $\Omega_{\mathcal{L}}$. So long as L and $\Omega_{\mathcal{L}}$ are small enough, C01b3a5$_S$ comparative likelihood relations are all representable with only certainty and impossibility. Weakening the restrictions on L and $\Omega_{\mathcal{L}}$, however, the discrepancy is obvious, and the correspondence fails. The restriction to finite sets operates in the same way. At most one of

$$\langle [0,1], 0, 1, \leq, + \rangle$$

$$\langle \mathbb{Q} \cap [0,1], 0, 1, \leq, + \rangle$$

$$\langle {}^*[0,1], 0, 1, \leq, + \rangle$$

is the correct likelihood structure, but the differences between them only manifest outside finite Ω.

As a result, each likelihood structure may either undergenerate or overgenerate comparative likelihood relations if the restriction to finite Ω is dropped. Supposing that rational comparative likelihood relations are all and only the C01b3a5$_S$ relations, for example, the probability structure undergenerates:

Proposition 3.4

Let L be a formal language with Boolean connectives and $\Omega_{\mathcal{L}}$ a model space for L. Then, there exists a binary relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ satisfying C01b3a5$_S$ which is not representable in $\langle [0,1], 0, 1, \leq, + \rangle$.

Indeed, this result generalizes; no single likelihood structure can represent C01b3a5$_S$ relations over arbitrary propositional languages.

Proposition 3.5

Let $\langle \mathbb{P}, \circ \rangle$ be a likelihood structure. Then, there exists a propositional language L, model space $\Omega_{\mathcal{L}}$, and C01b3a5$_S$ relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ such that $\precsim$ is not represented in $\langle \mathbb{P}, \circ \rangle$ by any NCI-likelihood assignment.

If C01b3a5$_S$ comparative likelihood relations are all rational, rational credences require an absolute ranking formalism which is not categorical.

The C01b3a5$_S$ representation argument is then doubly flawed. First, (P1) asserts that comparability is a requirement of rationality. This both contradicts a number of prima facie coherent accounts of rational credence and is not borne out by our earlier discussion of the relationship between rational credence and logic. Second, the conclusion of the C01b3a5$_S$ representation argument is overstated. In the finite case, C01b3a5$_S$ comparative likelihood relations are consistent with a number of non-isomorphic likelihood structures. Dropping the restriction to finite sets of possibilities, moreover, breaks the correspondence with the probability formalism. If rationality only requires a C01b3a5$_S$ comparative likelihood relation, then probabilism is false.

3.2.3 Representation Arguments from Comparative Likelihood

To circumvent the latter difficulty, a characterization of all rational comparative likelihood relations over arbitrary sets of possibilities together with a representation theorem establishing that these relations are exactly the relations representable with probability functions is required. Much of the formal work here is owed to Koopman (1940), Luce (1967), Krantz et al. (1971), Savage (1972), and Fine (1973). The strategy is to supplement the axioms for the finite case with both a decomposition axiom and an Archimedean axiom.

Decomposition Axioms

Decomposition axioms require that sets of possibilities can be partitioned into pieces of approximately equal size. The earliest proposal derives from de Finetti ([1937] 1980) and Koopman (1940) and concerns only the set of all possibilities Ω.

$\mathbf{CD_{DK}}$ For every $n \in \mathbb{N}^+$, there exists a partition $\{C_1, \ldots, C_n\}$ of Ω such that for any $i, j \in \{1, \ldots n\}$, $C_i \in \mathcal{F}$ and $C_i \sim C_j$.

Kraft et al. (1959) propose extending this to $\mathcal{F}$ as a whole.

Definition 3.3 Let $A \in \mathcal{F}$ and $\sim$ an equivalence relation over $\mathcal{F}$. A is *polarizable* in $\mathcal{F}$ if and only if there exists $A', A'' \in \mathcal{F}$ such that

- $A = A' \cup A''$
- $A' \cap A'' = \emptyset$
- $A' \sim A''$.

$\mathbf{CD_{KPS}}$ For any $A \in \mathcal{F}$, A is polarizable.

Drawing on Savage (1972), Fine (1973) works with a weaker condition that doesn't require equal likelihood between parts.

Definition 3.4 An *n-fold almost uniform partition of* Ω is a partition $\{C_1, \ldots, C_n\}$ of Ω such that if U_{k+1} is a union of $k+1$ sets from the partition and U_k is a union of k sets from the partition, $U_k \precsim U_{k+1}$

An n-fold almost uniform partition is thus a partition whose components, while not necessarily equal, are close enough that k-unions are always less likely than $k+1$ unions. The decomposition condition itself is,

$\mathbf{CD_F}$ For infinitely many $n \in \mathbb{N}^+$, there exists an n-fold almost uniform partition of Ω.

In the presence of C01b23a4, Savage (1972) proves that infinitely many n-fold almost uniform partitions entails n-fold almost uniform partitions for every $n \in \mathbb{N}^+$. (CD$_F$) thus guarantees n-fold almost uniform partitions generally.

All three of the decomposition axioms above force Ω to be infinite. Luce (1967)'s contribution does not:

CD$_\mathbf{L}$ If $A, B, C, D \in \mathcal{F}$ are such that $A \cap B = \emptyset$, $C \prec A$, and $D \precsim B$, then there exists $C', D', E \in \mathcal{F}$ such that:

(i) $E \sim A \cup B$

(ii) $C' \cap D' = \emptyset$

(iii) $C' \cup D' \subset E$

(iv) $C' \sim C$ and $D' \sim D$.

The most esoteric of the conditions, CD_L doesn't guarantee a straightforward decomposition of either a given set or Ω; rather, for any starting set expressible as the union of two disjoint sets and for any two sizes in our hierarchy smaller than our disjoint sets (at least one strictly so), we are guaranteed the existence of a correspondent for our original set which contains a disjoint set of each size.

Archimedean Axioms

Archimedean axioms ensure that the comparative likelihood relations under consideration do not violate one of the characteristic properties of the real numbers, viz. no infinitely large elements and no infinitely small elements. In a countable setting, all total orders are consistent with representation by real numbers:

Theorem 3.4 Cantor (1895)

If $\langle A, \preceq \rangle$ is a countable total order, then there exists an embedding of $\langle A, \preceq \rangle$ into $\langle \mathbb{R}, \leq \rangle$.

As the previous subsection showed, however, this not true in general. In the case of total orders, Debreu (1954) proved that real-valued representation coincides with the existence of a countable, order-dense subset.

Definition 3.5 Given a total order $\langle A, \preceq \rangle$ and $B \subseteq A$, B is *order-dense* in A if and only if for all $a, c \in A - B$ such that $c \prec a$, there exists $b \in B$ with $c \preceq b \preceq a$.

Theorem 3.5 Debreu (1954)

Let $\langle A, \leq\rangle$ be a total order. Then, the following are equivalent:

(i) There is a countable (possibly finite) order-dense subset of A.

(ii) There is an embedding of $\langle A, \leq\rangle$ into $\langle \mathbb{R}, \leq\rangle$.

The most straightforward means of ensuring that comparative likelihood relations are also representable in the real numbers is to directly impose Debreu's characterization of $\langle \mathbb{R}, \leq\rangle$.

CA$_\mathbf{D}$ There exists a countable, order-dense (relative to the strict total order induced by $\precsim$) $A \subseteq \mathcal{F}$.

This immediately guarantees representability of the comparative likelihood relation by some NCI-likelihood assignment into the unit interval.

Theorem 3.6 Fine (1973)

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation on $\mathcal{F}$. $\precsim$ satisfies C1b23a4A$_D$ if and only if there exists an NCI-likelihood assignment $\mu : \mathcal{F} \to [0, 1]$ which represents $\precsim$.

Given only C01b23a4A_D, however, the additivity requirement in probability functions remains problematic.

An alternative strategy is to cast the prohibition against infinitely large and infinitely small elements more directly in terms of the inconsistent disjunction operator $\circ$.

Definition 3.6 For a likelihood structure $\langle \mathbb{P}, \circ\rangle$ and $p \in \mathbb{P}$, a *standard sequence relative to* p is a (possibly finite) sequence of the form

$$p, \quad p \circ p, \quad p \circ p \circ p, \quad p \circ p \circ p \circ p, \quad \ldots$$

where each term is defined.

Stipulating that standard sequences are finite for every non-$\dot{0}$ likelihood p thus naturally blocks infinitely small elements. Unfortunately, the purely comparative analogue of a standard sequence is significantly more complicated.

Definition 3.7 Let $\mathcal{F}$ be an algebra, $\sim$ an equivalence relation on $\mathcal{F}$, and $A \in \mathcal{F}$. A sequence of sets $A_1, A_2, \ldots$ drawn from $\mathcal{F}$ is a *standard sequence relative to A* if and only if for any A_i, there exists $B_i, C_i \in \mathcal{F}$ such that:

(i) $A_1 = B_1$ and $B_1 \sim A$

(ii) $B_i \cap C_i = \emptyset$

(iii) $B_i \sim A_i$

(iv) $C_i \sim A$

(v) $A_{i+1} = B_i \cup C_i$.

The focus here is not on re-adding A to our sequence since set union does not recognize multiplicity; rather, each A_i is some set from $\mathcal{F}$ with the likelihood we would expect from unioning copies of A together. Each C_i is thus a copy of A (i.e., a set with equivalent likelihood) while each B_i is a disjoint set whose likelihood is equal to that of the previous stage in our sequence, thus allowing the union $B_i \cup C_i$ to represent the "addition" of a new copy.

The actual axiom we need to impose is then

CA$_\mathbf{S}$. For every $A \in \mathcal{F}_\mathcal{L}$ with $\emptyset \prec A$, every standard sequence relative to A is finite.

Setting aside sets equivalent to the empty set, no standard sequence for a set is unbounded, and thus no set can force an infinitely small value by way of additivity.

Dual Purpose Axioms

Unlike other commentators, Savage (1972) introduces an axiom which acts simultaneously as a tool for decomposition and rules out both infinitely small and infinitely large elements.

CS. If $A, B \in \mathcal{F}$ and $B \prec A$, then there exists a partition $\{C_1, \ldots, C_n\}$ of Ω such that $C_i \in \mathcal{F}$ and $B \cup C_i \prec A$ for every i.

Savage also provides a useful alternative characterization of (CS).

Definition 3.8 For $A, B \in \mathcal{F}$, A is *almost equivalent* to B—notated $A \sim^* B$—if and only if for any $C, D \in \mathcal{F}$ with $\emptyset \prec C, D$ and $A \cap C = B \cap D = \emptyset$, $B \precsim A \cup C$ and $A \precsim B \cup D$.

Informally, A and B are almost equivalent if and only if any disjoint set in $\mathcal{F}$ is enough to reverse the comparative likelihood relationship between the two. (CS) can now be expressed as the conjunction of two more intuitive properties.

Definition 3.9 A binary relation $\precsim$ on an algebra $\mathcal{F}$ is *fine* if and only if for every $A \in \mathcal{F}$ with $\emptyset \prec A$, there exists a partition $\{C_1, \ldots, C_n\}$ of Ω such that $C_i \precsim A$ for $i = 1, \ldots, n$.

Definition 3.10 A binary relation $\precsim$ on an algebra $\mathcal{F}$ is *tight* if and only if for every $A, B \in \mathcal{F}$, if $A \sim^* B$, then $A \sim B$.

Theorem 3.7 Savage (1972)

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation on $\mathcal{F}$ satisfying C01b23b4. $\precsim$ satisfies CD_S if and only if $\precsim$ is both fine and tight.

Several implications between Savage's (CS) and the other axioms detailed in this section are known to either hold or fail. In general, (CS) is a relatively powerful decomposition condition.

Proposition 3.6

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation satisfying C01b23b4. Then,

(i) Savage (1972): (CS) $\Rightarrow$ (CD_{DK}) and (CD_{DK}) $\not\Rightarrow$ (CS)

(ii) Luce (1967): (CS) $\Rightarrow$ (CD_L) and (CD_L $\not\Rightarrow$ (CS).

Since (CD_{DK}) implies (CD_F), Savage's proposal is strictly stronger than the decomposition requirements imposed by Luce and Fine.

Representation Theorems and Arguments for Finitely Additive Probabilism

Three different combinations of the axioms above are known to suffice for representation by finitely additive probability functions.[8]

Theorem 3.8 Savage (1972)

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation on $\mathcal{F}$. If $\precsim$ satisfies C1b23a4S, then there exists a unique finitely additive probability function $\mu : \mathcal{F} \to [0, 1]$ which represents $\precsim$.

Theorem 3.9 Fine (1973)

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation on $\mathcal{F}$. If $\precsim$ satisfies C1b23a4$D_F A_D$, then there exists a unique finitely additive probability function $\mu : \mathcal{F} \to [0, 1]$ which represents $\precsim$.

[8]A *finitely additive probability function* over an algebra $\mathcal{F}$ is a function $\mu : \mathcal{F} \to \mathbb{R}$ which satisfies (K1), (K2), and additivity over finite collections of pairwise disjoint events. Every probability functions is thus a finitely additive probability function, but not every finitely additive probability function is a probability function.

Theorem 3.10 Luce (1967); Krantz et al. (1971)

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation on $\mathcal{F}$ satisfying C01b23a4D$_L$A$_S$. Then, there exists a unique, finitely additive probability function $\mu : \mathcal{F} \to [0, 1]$ which represents $\precsim$.

This in turn gives rise to three representation arguments for finitely additive probabilism. I combine them here for simplicity.

C012B3A4+S/D$_F$A$_D$/D$_L$A$_S$ REPRESENTATION ARGUMENT

(P1) (C0), (C1b), (C2), (C3a), and (C4) are rational restrictions on comparative likelihood relations.

(P2) Either (CS), (CD$_F$)∧(CA$_D$), or (CD$_L$)∧(CA$_S$) is a rational restriction on comparative likelihood relations.

(P3) If a comparative likelihood relation satisfies one of

C01b3a4S

C01b3a4D$_F$A$_D$

C01b3a4D$_L$A$_S$

then it is representable by a unique finitely additive probability function μ : $\mathcal{F} \to [0, 1]$.

(C) If c is a rational credence over an algebra $\mathcal{F}$, then c is a finitely additive probability function.

Representation Theorems and Arguments for Probabilism

In all three cases the jump to a probability function can be made with one final continuity axiom.

C8. For any countable collection of downward nested[9] sets $\{B_i\}$, if $\cap_i B_i = \emptyset$, then for any set $\{A_i : \emptyset \prec A_i \preceq B_i\}$

$$\cap_i A_i = \emptyset.$$

[9]A countable collection of sets $\{B_i\}_{i \in I}$ is *downward nested* if and only if $B_{i+1} \subseteq B_i$ for all i.

Theorem 3.11 Fine (1973)

Let Ω be a non-empty set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation on $\mathcal{F}$. If $\precsim$ is representable by a finitely additive probability function $\mu : \mathcal{F} \to [0, 1]$, then μ is countably additive if and only if $\precsim$ satisfies (C8).

A final representation argument from comparative likelihood to probabilism is thus available.

C01B23A48+S/D$_F$A$_D$/D$_L$A$_S$ REPRESENTATION ARGUMENT

(P1) (C0), (C1b), (C2), (C3a), and (C4) are rational restrictions on comparative likelihood relations.

(P2) Either (CS), (CD$_F$)∧(CA$_D$), or (CD$_L$)∧(CA$_S$) is a rational restriction on comparative likelihood relations.

(P3) (C8) is a rational restriction on comparative likelihood relations.

(P4) If a comparative likelihood relation satisfies one of

C01b23a4S8

C01b23a48D$_F$A$_D$8

C01b23a4D$_L$A$_S$8

then it is representable by a unique probability function $\mu : \mathcal{F} \to [0, 1]$.

(C) If c is a rational credence over an algebra $\mathcal{F}$, then c is a probability function.

At last, representation by a probability function has been guaranteed.

3.2.4 The Problem with Arguments from Comparative Likelihood

Structurally, both the C012b3a4+S/D$_F$A$_D$/D$_L$A$_S$ representation argument for finitely additive probabilism and the C01b23a48+S/D$_F$A$_D$/D$_L$A$_S$ representation argument for probabilism do not fit the outline of the ideal representation argument. In both cases, the final premise as well as the underlying theorems supply sufficient but not necessary conditions for representation. This is a direct result of the decomposition axioms requiring a minimal amount of structure in the algebras themselves. All things considered, this discrepancy is not too worrying; each of (CD$_{DK}$), (CD$_{KPS}$), (CD$_F$), and (CD$_L$) are intuitively true in a natural extension of any given probability space.[10] As Savage (1972) advises, we need only find a coin and start flipping it.

The real problem with representation arguments from comparative likelihood is the choice of axioms. The probabilist was already forced to beg the question in the finite case, and the situation has only worsened. Setting aside the issue of comparability, (CA_D), (CA_S), and (CS) are all designed to force the Archimedean property. Not only does this too go beyond the purely logical restrictions explored in Chapter 2, it flies directly in the face of several robust intuitions about comparative likelihood. If we are willing to allow infinite sets of possibilities, is it not rationally permissible to judge each single possibility equally likely? (CA_S) and (CS) entail that it is not; (CA_D) together with (CD_F) entails that it is not. Is it not rationally permissible to judge each single possibility more likely than a contradiction? Either (CS) or (CA_S) together with (CD_L) entail that it is not; (CA_D) together with (CD_F) entails that it is not. The motivation for these answers is not any principle of rationality but rather that these answers are necessary for representation in terms of a probability function. Representation arguments from comparative likelihood fail because the axioms they must supply are implausible as purely rational constraints. Probabilism gains no ground here.

3.3 Gradational Accuracy Arguments

Gradational accuracy arguments for probabilism derive from Rosenkrantz (1981) and Joyce (1998) though inspiration is drawn also from work on calibration (van Fraassen 1983; Shimony 1988). The explicit goal of Joyce (1998) is to identify a nonpragmatic criterion of success for credences, a means of separating "good" credences from "poor" without reference to their material efficacy (cf. the Dutch book argument). Joyce proposes accuracy for this role, arguing that credences are successful in so far as they are accurate and unsuccessful in so far as they are not. We will not contest this; the current section will instead focus only on whether or not accuracy suffices to establish probabilism.

Joyce (1998) motivates his accuracy-based approach to rational credences with an analogy to traditional epistemology. Joyce holds that in the case of full belief there is a clear and uncontroversial notion of epistemic success:

> An epistemically rational agent must strive to hold a system of full beliefs that strikes the best attainable overall balance between the epistemic good of believing truths and the epistemic evil of believing falsehoods (fully believing a truth is better than being agnostic, being agnostic is better than fully believing a falsehood). (577)

Drawing inspiration from Jeffrey (1986), Joyce contends that partial beliefs or credences admit of a similar norm:

[10] In truth, I think there is good reason to reject this intuition and thus these decomposition axioms. The current argument can, however, be made independent of these concerns, and so I leave them aside here.

> An epistemically rational agent must evaluate partial beliefs on the basis of their gradational accuracy, and she must strive to hold a system of partial beliefs that, in her best judgment, is likely to have an overall level of gradational accuracy at least as high as that of any alternative system she might adopt. (579)

The difference between full belief and credences is thus like the difference between guessing a value and estimating it. The former is correct only when the guess matches reality exactly. The latter, however, are evaluated on a closeness criteria—the goal isn't to get the correct value but rather to be as close as possible to the correct value.

3.3.1 Forecasts and Scoring Rules

While philosophically novel, accuracy arguments for probabilism co-opt a preexisting formal framework for scoring forecasts (de Finetti 1979; Savage 1971; Lindley 1982; Predd et al. 2009). Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\mathcal{E}$ an n-tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$. A *forecast* f for $\mathcal{E}$ is an element of $[0, 1]^n$. A *scoring rule* is a function from a forecast f and particular outcome $w \in \Omega$ to a score in $[0, \infty]$. The most well-known example of a scoring rule is the Brier score:

$$S_B(f, w) = \sum_i \frac{1}{n}[w(E_i) - f_i]^2$$

where

$$w(E_i) = \begin{cases} 1 & w \in E_i \\ 0 & w \notin E_i. \end{cases}$$

A straightforward generalization of the Brier score gives the set of quadratic loss rules, scoring rules of the form

$$S(f, w) = \sum_i \lambda_i [w(E_i) - f_i]^2$$

where $\lambda_i \in \mathbb{R}^+ - \{0\}$ for every i and $\sum_i \lambda_i = 1$. Note that the Brier score and quadratic loss rules assign larger values to worse forecasts. For this reason, scoring rules are often (and profitably) thought of as penalties.

The foundational theorems in this area are owed to both de Finetti (1979) and Savage (1971) and relate forecasts scored by quadratic loss rules to probabilities:

Theorem 3.12 **de Finetti (1979); Savage (1971)**

Let Ω be a finite set of possibilities, $\mathcal{F}$ an algebra over Ω, S a quadratic loss rule, and $\mathcal{E}$ a tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$.

(i) For any forecast f over $\mathcal{E}$, if f is not probabilistic, then there exists a probabilistic forecast f^* such that $S(f^*, w) < S(f, w)$ for every $w \in \Omega$.

(ii) For any probabilistic forecast f over $\mathcal{E}$, there exists no forecast f^* such that $S(f^*, w) \leq S(f, w)$ for every $w \in \Omega$ and $S(f^*, w) < S(f, w)$ for some $w \in \Omega$.

Theorem 3.13 **de Finetti (1979); Savage (1971)**

Let Ω be a finite set of possibilities, $\mathcal{F}$ an algebra over Ω, S a quadratic loss rule, and $\mathcal{E}$ a tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$. For any probability function $\mu : \mathcal{F} \to [0, 1]$, setting $f = \langle \mu(E_1), \ldots, \mu(E_n) \rangle$ minimizes expected score with respect to μ:

$$\sum_j \mu(w_j) S(f, w_j).$$

The first of these theorems shows that probabilistic forecasts are better than non-probabilistic forecasts with respect to quadratic loss rules. The second establishes that, given a probability function, the forecast which matches this function has the minimal expected score. Every probability function is, by its own lights, best.

These two ideas are brought together in a later result due to Predd et al. (2009).

Definition 3.11 A scoring rule $S(f, w) = \sum_i s(f_i, w(E_i))$ for $s : [0, 1] \times \{0, 1\} \to [0, \infty]$ is *proper* if and only if

- $(p)s(f_i, 1) + (1 - p)s(f_i, 0)$ is uniquely minimized at $f_i = p$ for all $p \in [0, 1]$, and
- s is continuous, i.e., for $i \in \{0, 1\}$,

$$\lim_{n \to \infty} s(x_n, i) = s(x, i)$$

for any sequence $x_n \in [0, 1]$ converging to x.

Proper scoring rules are thus continuous functions which reward honesty; the f_i value which does best by the lights of p is p itself.[11] It is easy to verify that the Brier score is an example of a proper scoring rule.

[11] Best, that is, with respect to the expression $(p)s(f_i, 1) + (1 - p)s(f_i, 0)$. While this expression enforces normalization with the use of $(1 - p)$ in the second term, something like this is required if we are to have any scoring rule which meets the criteria.

Theorem 3.14 **Predd et al. (2009)**

Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, S a proper scoring rule, and $\mathcal{E}$ a tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$.

(i) For any forecast f over $\mathcal{E}$, if f is not probabilistic, then there exists a probabilistic forecast f^* such that $S(f^*, w) < S(f, w)$ for every $w \in \Omega$.

(ii) For any probabilistic forecast f over $\mathcal{E}$, there exists no forecast $f^* \neq f$ such that $S(f^*, w) \leq S(f, w)$ for every $w \in \Omega$.

Proper scoring rules in general thus hold that probabilistic forecasts are better than non-probabilistic forecasts.

3.3.2 Three Accuracy Arguments for Probabilism

Accuracy arguments for probabilism now fall out of these formal results by interpreting forecasts as credences and presenting scoring rules as measures of accuracy. Rosenkrantz (1981), for example, sketches something like the following argument:

ROSENKRANTZ (1981) ACCURACY ARGUMENT FOR PROBABILISM

($P1_R$) Any reasonable measure of accuracy for credences is a proper scoring rule.

($P2_R$) For any proper scoring rule and any non-probabilistic credence f, there exists a probabilistic credence f^* which scores strictly better than f no matter how the world turns out.

($P3_R$) It is irrational to hold a credence f which is strictly less accurate than another credence f^* no matter how the world turns out.

(C) If f is a rational credence, then f is a probability function.

Taking each premise in order, ($P1_R$) ought to be accepted because reasonable measures of accuracy ought not encourage even non-probabilistic forecasters "to misrepresent, at any rate, his actual *relative* degrees of belief" (Rosenkrantz 1981, 2.2-3). Credences—even non-probabilistic ones—ought to be maximally accurate by their own lights. While only familiar with de Finetti (1979), Rosenkrantz (1981) conjectured that something like the Predd et al. (2009) theorem held.[12] From the modern perspective, ($P2_R$) is true simply in virtue of this result. Finally, ($P3_R$) encodes only that accuracy is the criteria of epistemic

[12] Rosenkrantz (1981) specifically conjectured that the quadratic loss rules were the only scoring rules

success for credences. Putting all three together, we will have "nail[ed] down the sense in which incoherence incurs a purely cognitive penalty" (Rosenkrantz 1981, 2.2-3).

While Joyce (1998) agrees that this argument favors probabilism, he contends that it fails to do so from neutral premises. In particular, the use of expectation built into the definition of a proper scoring rule is only motivated if credences are, in fact, probabilities. If credences are non-probabilistic, it unclear why we would desire that $f_i = p$ is a unique minimum for $(p)s(f_i, 1) + (1 - p)s(f_i, 0)$; this expression has lost its significance. Joyce himself endeavors to give a more satisfactory result by replacing the requirement that scoring rules be proper with a set of six axioms for accuracy measures which he takes to be both neutral and intuitive.

Let Ω be a set of possibilities and $\mathcal{F}$ a countable algebra over Ω. Joyce (1998) takes accuracy measures I to be functions

$$I : B \times W \to [0, \infty]$$

where B is the set of all credence functions over $\mathcal{F}$.[13] Joyce then adopts the following six axioms for I:

(a) *Structure*
For every $w \in \Omega$, $I(b, w)$ is a non-negative and continuous function of b such that

$$\lim_{b(E)\to\infty} I(b, w) = \infty$$

for any $E \in \mathcal{F}$.

(b) *Extensionality*
At every w, $I(b, w)$ is a function of only the values assigned by w and b to $E \in \mathcal{F}$.

(c) *Dominance*
If $b(E) = b'(E)$ for every $E \in \mathcal{F}$ other than E^*, then $I(b, w) > I(b', w)$ if and only if $|w(E^*) - b(E^*)| > |w(E^*) - b'(E^*)|$.

(d) *Normality*
If $|w(E) - b(E)| = |w'(E) - b'(E)|$ for all $E \in \mathcal{F}$, then $I(b, w) = I(b', w')$.

(e) *Weak Convexity*
If $I(b, w) = I(b', w)$, then $I(b, w) \geq I(\frac{1}{2}b + \frac{1}{2}b', w)$ with identity only if $b = b'$.

---

which (i) decrease with proximity to the true value and increase with proximity to the false value, (ii) are twice differentiable, and (iii) minimized expected score for normalized $\mu : \mathcal{F} \to [0, 1]$ over a finite $\mathcal{F}$. If correct, the de Finetti (1979)-Savage (1971) theorem would then establish probabilistic dominance.

[13] Joyce (1998) never explicitly specifies the exact bounds on this space. Given his axioms, he appears to intend that B contains at least all functions $b : \mathcal{F} \to \mathbb{R}^+$. In his criticism, Maher (2002) explicitly takes B to be the set of all $b : \mathcal{F} \to \mathbb{R}$. By Joyce (2009), Joyce's formal framework has changed. $\mathcal{F}$ is stipulated to be finite, and B is defined as the set of all functions $b : \mathcal{F} \to [0, 1]$.

(f) *Symmetry*
If $I(b, w) = I(b', w)$, then for any $\lambda \in [0, 1]$, $I(\lambda b + (1-\lambda)b', w) = I((1-\lambda)b + \lambda b', w)$.

Structure guarantees that accuracy is non-negative, that small changes in credences result in small changes in accuracy, and that inaccuracy increases without limit as credences do. *Extensionality* requires that nothing besides the facts and an agent's credence matter for inaccuracy. *Dominance* ensures that accuracy increases with an agent's confidence in a truth and a decreases with an agent's confidence in a falsehood. *Normality* dictates that credences equally far from the truth of every $E \in \mathcal{F}$ are equally accurate. *Weak convexity* encodes the maxim that extremism is not a virtue. If a particular shift to an agent's credences doesn't improve accuracy, then doing it twice won't improve accuracy either. Finally, *symmetry* requires that, given equally inaccurate starting credences b and b', accuracy gains occur equally when shifting b a little towards b' and b' a little towards b.

From these six axioms Joyce establishes a dominance result for probabilistic credences.

Theorem 3.15 Joyce (1998)

Let Ω be a set of possibilities, $\mathcal{F}$ a countable algebra over Ω, and I an accuracy measure satisfying axioms (a)-(f). Then,

(i) for any $b : \mathcal{F} \to \mathbb{R}$, if b is not probabilistic, then there exists a probabilistic credence b^* such that $I(b^*, w) < I(b, w)$ for every $w \in \Omega$.

This generates the following argument for probabilism:

Joyce (1998) Accuracy Argument for Probabilism

(P1$_J$) Any reasonable measure of accuracy for credences satisfies axioms (a)-(f).

(P2$_J$) For any accuracy measure I satisfying (a)-(f) and any non-probabilistic credence b, there exists a probabilistic credence b^* which is strictly more accurate than b no matter how the world turns out.

(P3$_J$) It is irrational to hold a credence b which is strictly less accurate than another credence b^* no matter how the world turns out.

(C) If b is a rational credence, then b is a probability function.

Since Joyce's accuracy axioms were selected precisely because they are obviously true, (P1$_J$) is true. (P2$_J$) meanwhile is simply Joyce's theorem. Finally, (P3$_J$) is guaranteed if we take accuracy as the criterion of epistemic success for credences. Put together, (P1$_J$), (P2$_J$), and (P3$_J$) deliver probabilism.

Since its publication, three salient lines of objection to Joyce (1998)'s argument have appeared. First, Joyce's theorem provides no guarantee that the more accurate credence f^* does not change depending on the particular accuracy measure I used (Hájek 2009). It would be significantly better to prove that for any non-probabilistic credence f there exists a probabilistic credence f^* such that for any accuracy measure I, f^* is strictly more accurate than f no matter how the world turns out. Given only Joyce's weaker result, (P3$_J$) must be understood as claiming that f is irrational if it is accuracy-dominated according to any reasonable measure of accuracy (rather than all). Second, it is not obvious that accuracy-dominated credences are irrational; these credences may have a virtue other than accuracy, perhaps according with your evidence (Easwaran and Fitelson 2012) or determining the truth of particular propositions (Caie 2013; Greaves 2013). Third and finally, Maher (2002) challenges Joyce's *weak convexity* and *symmetry* axioms by noting that several intuitive accuracy measures violate both, most notably the distance measure $I(b, w) = \sum_i |b(E_i) - w(E_i)|$. Maher (2002) further proves that Joyce's theorem fails without these axioms.

In response to these objections, Pettigrew (2016) presents our third and final accuracy argument for probabilism:

> Pettigrew (2016) Accuracy Argument for Probabilism
>
> (P1$_P$) The only reasonable measure of accuracy for credences is the Brier score (or a linear transformation thereof).
>
> (P2$_P$) For any non-probabilistic credence f, there exists a probabilistic credence f^* with a strictly better Brier score than f no matter how the world turns out.
>
> (P3$_P$) It is irrational to hold a credence f given that there exists another credence f^* such that (i) f is strictly less accurate than f^* in every world and (ii) there is no credence f' that is at least as accurate as f^* in every world and strictly more accurate in some particular world.
>
> ---
>
> (C) If f is a rational credence, then f is a probability function.

Like Joyce (1998), Pettigrew's approach to measuring accuracy is axiomatic. Unlike Joyce (1998), the axioms endorsed by Pettigrew (2016) suffice to restrict accuracy measures to the Brier score or some linear transformation thereof. For the sake of brevity, I omit the axioms here; all are, however, similar in style to those introduced by Joyce. These axioms, then, are to ground (P1$_P$). (P2$_P$) is, as usual, secured by a formal theorem; in this case, the original de Finetti (1979)-Savage (1971) result. Given this theorem, our final premise (P3$_P$) can be weakened to require an alternative credence f^* which both strictly accuracy dominates f and is not itself even weakly accuracy dominated. Combining (P1$_P$), (P2$_P$), and (P3$_P$), we again have probabilism.

3.3.3 The Problem with Accuracy Arguments

The first and second premises in all three accuracy arguments are false. Further, the underlying error is by this point quite familiar. Each of ($P1_R$), ($P1_J$), and ($P1_P$) tacitly assume that credences are real-valued. If credences are not real-valued, the definition of a proper scoring rule, the Brier score, Joyce (1998)'s axioms, and Pettigrew (2016)'s axioms are underdetermined. The situation is even more damning with ($P2_R$), ($P2_J$), and ($P2_P$). All three are supposed to be secured by a corresponding formal result. These formal results, however, are about *real-valued functions* not arbitrary credences as each of ($P2_R$), ($P2_J$), and ($P2_P$) claim. Perhaps the mental life of accuracy proponents is significantly different from mine, but my credences don't appear to be so sharp as to be real-valued nor do they appear to satisfy either comparability or the Archimedean property. Here we have arguments for the conditional conclusion 'if credences are real-valued, then probabilism' masquerading as arguments for probabilism.

A simple example helps to drive this point home. Consider the likelihood space with domain $[0,1]\times[0,1]$, minimal element $\langle 0,0\rangle$, maximal element $\langle 1,1\rangle$, and the lexicographic ordering

$$\langle x_1, y_1\rangle \leq \langle x_2, y_2\rangle \text{ if and only if either } x_1 < x_2 \text{ or both } x_1 = x_2 \text{ and } y_1 \leq y_2.$$

This space naturally lifts into a likelihood structure by setting

$$\langle x_1, y_1\rangle \circ \langle x_2, y_2\rangle = \begin{cases} \langle x_1 + x_2, y_1 + y_2\rangle & \text{if } x_1 + x_2, y_1 + y_2 \leq 1 \\ \uparrow & \text{otherwise.} \end{cases}$$

While reminiscent of the conventional probability structure $\langle [0,1], \leq, +\rangle$, the behavior of the $\circ$-operation makes this 2-probability structure distinct.[14] Finally, we can obtain the 2-probability functions—the analogue of probability functions—by taking all NCI-likelihood assignments which conform to $\circ$ in even the case of a countably infinite collection of disjoint sets. This is equivalent to saying that a function $\mu : \mathcal{F} \to [0,1]\times[0,1]$ is a 2-probability function if and only if μ satisfies

2K1. $\mu(\emptyset) = \langle 0,0\rangle$;

2K2. $\mu(\Omega) = \langle 1,0\rangle$;

2K3. For any countable collection of disjoint sets $\{E_i\} \subseteq \mathcal{F}$,

$$\mu(\cup\{E_i\}) = \langle \sum_i \mu(E_i)[1], \sum_i \mu(E_i)[2] \,\rangle.$$

[14]Both structures contain a unique element p such that $p \circ p = \dot{1}$, $\frac{1}{2}$ in the probability structure and $\langle \frac{1}{2}, \frac{1}{2}\rangle$ in the 2-probability structure. Any isomorphism between the two structures must thus take $\frac{1}{2}$ to $\langle \frac{1}{2}, \frac{1}{2}\rangle$. Note next that any element in the probability structure strictly below $\frac{1}{2}$ can be "added" to itself at least once. $\langle 0,1\rangle$, however, is an element strictly below $\langle \frac{1}{2}, \frac{1}{2}\rangle$ which cannot be "added" to itself. It follows immediately that no isomorphism exists between the probability structure and the 2-probability structure.

where $\mu(E_i)[j]$ is the j entry in $\mu(E_i)$.

Just as there is a connection between forecasts, the Brier score, and probability functions, there is a parallel relationship between 2-forecasts, the 2-Brier score, and 2-probability functions. Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\mathcal{E}$ an n-tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$. A *2-forecast* f for $\mathcal{E}$ is an n-tuple of elements from $[0,1]^2$. A *2-scoring rule* $S(f, w)$ is a function from a 2-forecast f and particular outcome $w \in \Omega$ to a score in $\mathbb{R}^+ \times \mathbb{R}^+$. The most natural example of a 2-scoring rule is the 2-Brier score:

$$S_B(f, w) = \langle \sum_i \frac{1}{n}[w(E_i)[1] - f_i[1]]^2, \sum_i \frac{1}{n}[w(E_i)[2] - f_i[2]]^2 \rangle$$

where

$$w(E_i)[j] = \begin{cases} 1 & w \in E_i \\ 0 & w \notin E_i \end{cases}$$

and $f_i[j]$ is the jth entry in the ith value of f. If $f = \langle\langle 1, 0\rangle, \langle \frac{1}{2}, 0\rangle\rangle$ for example, then $f_1[1] = 1$, $f_1[2] = 0$, $f_2[1] = \frac{1}{2}$, and $f_2[2] = 0$. The 2-Brier score itself simply returns a pair containing the Brier score with respect to the first entry in f_i and the Brier score with respect to the second entry in f_i. Just as with the 2-probability structure, these values from $\mathbb{R}^+ \times \mathbb{R}^+$ are ordered lexicographically by $\preceq$. A score of $\langle 3, 7.9\rangle$ is strictly less than a score of $\langle 5.1, 1\rangle$, notated $\langle 3, 7.9\rangle \prec \langle 5.1, 1\rangle$.

An analogue of the de Finetti (1979)-Savage (1971) result for 2-forecasts, the 2-Brier score, and the 2-probability functions can be easily established:

Proposition 3.7

Let Ω be a finite set of possibilities, $\mathcal{F}$ an algebra over Ω, S_B the 2-Brier score, and $\mathcal{E}$ a tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$.

(i) For any 2-forecast f over $\mathcal{E}$, if f is not 2-probabilistic, then there exists a 2-probabilistic 2-forecast f^* such that $S_B(f^*, w) \prec S_B(f, w)$ for every $w \in \Omega$.

(ii) For any 2-probabilistic 2-forecast f over $\mathcal{E}$, there exists no 2-forecast f^* such that $S(f^*, w) \preceq S(f, w)$ for every $w \in \Omega$ and $S(f^*, w) \prec S(f, w)$ for some $w \in \Omega$.

Theorem in hand, we now have an argument for 2-probabilism:

Accuracy Argument for 2-Probabilism

(P1) The only reasonable measure of accuracy for credences is the 2-Brier score.

(P2) For any credence f that is not 2-probabilistic, there exists a 2-probabilistic credence f^* with a strictly better 2-Brier score than f no matter how the world turns out.

(P3) It is irrational to hold a credence f given that there exists another credence f^* such that (i) f is strictly less accurate than f^* in every world and (ii) there is no credence f' that is at least as accurate as f^* in every world and strictly more accurate in some particular world.

(C) If an agent is rational, then their credence c over $\mathcal{F}$ is a 2-probability function.

If the accuracy arguments provided by Rosenkrantz (1981), Joyce (1998), and Pettigrew (2016) establish probabilism, this accuracy argument establishes 2-probabilism. Indeed, it is not difficult to generalize this example into an argument for k-probabilism for any $k \in \mathbb{N} - \{0\}$.

This is plainly absurd. The accuracy argument for 2-probabilism fails because it takes the restriction to the 2-probability structure for granted. All three accuracy arguments do precisely the same thing only for the probability structure. The only point of disanalogy is that the probability structure is conventional, and so the restriction is liable to be overlooked in that case. A satisfactory argument for probabilism must establish that the probability structure is the correct likelihood structure for rational credence not simply assume it.

Part II

A Logical Theory of Confirmation

Chapter 4

Foundations of Confirmation

The origins of confirmation theory are typically traced back to the work of W.E. Johnson and two of his Cambridge students: Sir Harold Jeffreys and John Maynard Keynes (Gillies 2000; Hacking 2006). While Jeffreys (1967) regarded the probability formalism as a conventional and somewhat arbitrary choice for confirmation, Keynes (1921) actively rejected it as inadequate. Despite this early hostility, work on confirmation eventually came to focus exclusively on the probability formalism under the dual influences of Carnap (1962) and the subjective Bayesian account of rational credence. This emphasis on probability has continued unabated to date with ruinous consequences for the original program.

The current chapter sets the foundations for the proposed account of confirmation. The first section summarizes the contemporary case against confirmation. This case resolves into three different objections: Ramsey's skepticism, d'Alembert's riddle, and Bertrand's paradox. While the primary objective of both this chapter and the next is a satisfactory account of confirmation, these three objections supply a series of challenges that deeply inform the eventual proposal. Accordingly, the discussion is laid out in terms of these objections and the refinements they require in the proposed account.

The rest of the current chapter tackles the first two of these objections. Ramsey's skepticism is taken up in section 4.2 alongside a basic discussion of rankings and the relationship between classical logic and confirmation. Despite its obscurity, d'Alembert's riddle provides a more substantial challenge and takes up the final two sections. By the last of these, an adequate theory of confirmation for finite cases has been constructed and defended.

4.1 Three Objections

There are three objections which face any theory of confirmation and which have served to convince contemporary theorists that there is no such thing: Ramsey's skepticism, d'Alembert's riddle, and Bertrand's paradox. The first and most basic of these objections is the general skepticism famously expressed by Frank Ramsey in his posthumous Ramsey (1931). In his discussion of Keynes' early work on confirmation, Ramsey—widely recognized as a brilliant mathematician and logician—observed that the confirmation relations

Keynes purported to analyze were entirely inaccessible to him. Still further, these allegedly objective and *a priori* relations appeared equally inaccessible to anyone outside Keynes himself. For all the intuitive appeal of confirmation, there existed little agreement on even the simplest cases, and thus Ramsey argued that there is a *prima facie* case here for skepticism. For all its initial appeal, confirmation may be just "an ancient habit of thought that dies hard" (Howson 1991, 550).

This case for skepticism is bolstered by the second and third objections to confirmation. The second objection, d'Alembert's riddle, was first formulated in the 18th century by the French mathematician Jean-Baptiste le Rond d'Alembert. The central difficulty has reappeared several times since, most notably with the book paradox presented by Keynes (1921) and the solution space observations made by Edwards (1978). During d'Alembert's life, confidence in the power of probability was nearing its high-water mark, and theorists seemed on the verge of a successful account of confirmation. The implicit basis for this confidence was an inference latter dubbed 'the principle of non-sufficient reason'.

Theories of confirmation attempt to move from lack of evidence to degrees of confirmation. First used by Bernoulli (1713), the principle of non-sufficient reason is both the earliest inference of this kind and the most famous (Keynes 1921; Gillies 2000). Nearly three centuries after its first use, Keynes found this received name clumsy and provided its contemporary moniker: *the principle of indifference*.

> *The Principle of Indifference*: "If there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability" (Keynes 1921, 45).[1]

While theories of confirmation are not required to validate either this principle or some variant thereof, confirmation and (POI) are a natural pairing. (POI) itself is only plausible if the "probabilities" it mentions are both objective and "knowledge"-relative. Theories of confirmation meanwhile would be well served by a principle which forges an intuitive connection between evidence and likelihood. Accordingly, proponents of confirmation have universally defended some variation of (POI) and vice-versa for proponents of (POI).

For all this principle's intuitive appeal, d'Alembert's riddle displays a serious lacuna in its application. In order to apply (POI) properly, there must first exist a well-defined collection of alternatives. Still further, if (POI) is to be the basis for a theory of confirmation, then either this collection of alternatives should itself be objective or it must be established that

[1]As the locus classicus for the principle of indifference, Keynes' presentation has exerted significant influence on later work and is as authoritative a statement of the principle as is possible. For comparison, Jeffreys provides a similar and somewhat simpler gloss: "If there is no reason to believe one hypothesis rather than another, the probabilities are equal" (33). In the manner characteristic of this area, the principle of indifference is nevertheless somewhat amorphous in nature. Not only do different commentators often have conflicting commitments with respect to "probability", but the progenitors of the principle also used it only implicitly. Accordingly, precise formulations will be labeled as 'variants of the principle of indifference' even though it is entirely possible that particular authors intended a particular variant as *the* principle of indifference.

all acceptable choices here lead to the same confirmation relations. It is thus insufficient for a collection of alternatives to merely be supplied by intuition or general agreement; if (POI) is to underpin a theory of confirmation, a rigorous account of the alternatives mentioned by the principle is required. D'Alembert's riddle forces this issue by describing a simple game and then outlining two conflicting sets of alternatives which both cohere with the provided description. Without a satisfactory account of the alternatives underlying (POI) and thus resolution to d'Alembert's riddle, any confirmation theory based upon (POI) is underdetermined.

The final and most influential objection to confirmation is Bertrand's paradox (sometimes also 'the multiple partitions problem'). Popularized in 19th century mathematics circles by the French mathematician Joseph Bertrand, the paradox rose to broad philosophical prominence only after its reformulation in van Fraassen (1989) as 'the mystery cube factory'. Further instances of the same phenomenon include both the wine and water paradox from von Mises (1957) and the volume - specific volume paradox presented in von Kries (1886). At its core, Bertrand's paradox presents a well-defined space of possibilities where the confirmation judgments endorsed by (POI)-style reasoning change depending on how that space is presented. Extended discussion of Bertrand's paradox will be postponed until the next chapter.

Not only then do Ramsey's observations give us reason to doubt the existence of a confirmation relation but d'Alembert's riddle and Bertrand's paradox purport to undermine the few robust intuitions available. These three objections together with the lack of any plausible formalization of confirmation have convinced most commentators that there is in fact no such thing (van Fraassen 1989; Howson 1991; Gillies 2000). The next two chapters construct a comparative theory of confirmation, addressing each of these objection along the way.

4.2 Ramsey's Skepticism

Ramsey (1931) famously sets out the foundations for the subjective Bayesian account of rational credence. Before giving his own positive proposal, however, Ramsey lodges a number of complaints against the theory of confirmation set out by Keynes (1921). Amongst these complaints, only a single objection is brought against confirmation itself:

> ...there really do not seem to be any such things as the [confirmation] relations [J.M. Keynes] describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. I do not perceive them, and if I am to be persuaded that they exist it must be by argument; moreover I shrewdly suspect that others do not perceive them either, because they are able to come to so very little agreement as to which of them relates any two given propositions. (Ramsey 1931, 161)

Ramsey here makes a simple argument for skepticism. If there is such a thing as confirmation, then there ought to be cases where these relations can be recognized by suitably careful and intelligent persons. There is little agreement on cases despite the existence of a large number of careful and intelligent commentators. Thus, there is no such thing as confirmation.

The most striking feature of Ramsey's objection to confirmation is that it is entirely negative. The lack of an accepted set of confirmation relations is the sole justification supplied for denying the existence of confirmation itself. Accordingly, Ramsey's skepticism only survives so long as the stock of accepted confirmation relations remains small. The remainder of this section rectifies this shortcoming. Contra Ramsey, intuitive instances of confirmation abound; the only scandal here is that these cases have gone unrecognized for so long.

The central difficulty with confirmation is the treatment of those propositions which outrun the evidence. Given a set of evidence E, classical logic divides propositions into three classes. First, those propositions which contradict the evidence, claims which are necessarily false given the truth of E. Second, those propositions which are entailed by the evidence, claims which are necessarily true given the truth of E. Third and finally, those propositions which neither contradict nor are entailed by the evidence, claims whose truth and falsity are consistent with E. The first and second categories supply a lower and upper extreme for confirmation, maximal dis-confirmation and maximal confirmation respectively. The true difficulty with confirmation is thus only in the ranking of contingent propositions.

It follows immediately that there exists a large stock of accepted confirmation relations. Whenever a claim P is entailed by evidence E, it follows both that P is maximally confirmed by E and that $\neg P$ is maximally disconfirmed by E. There is no lack here in sheer number of cases. Nevertheless, this stock of accepted confirmation relations does fall short in diversity. Ramsey's objective is better framed then not as a matter of quantity but of variety. Since confirmation differs from logical consequence in admitting multiple degrees, proponents of confirmation ought to be able to supply a number of intuitive confirmation relations which also vary in degree.

Even here, classical logic suffices to undercut Ramsey's objection. Viewed correctly, both propositional and first-order logic induce minimalist confirmation theories and therefore supply an infinite stock of intuitive confirmation rankings. There is thus no lack in intuitive confirmation relations, and Ramsey's skepticism about confirmation is entirely undeserved. Keynes, Ramsey, and subsequent commentators were simply focused on the wrong sort of ranking.

4.2.1 Minimal Comparative Confirmation

As a preliminary to the main proposal and rebuttal to Ramsey's skepticism, both propositional and first-order logic straightforwardly induce corresponding comparative confirmation relations over propositional and first-order languages respectively. Classical logic as a whole is, in a precise sense, just a minimalist comparative theory of confirmation.

Propositional Logic as Confirmation Theory

Let L be a propositional language, and let $\models_{\mathrm{PL}}$ denote the propositional consequence relation. For consistent $\Sigma \subseteq L$ and any $\varphi, \psi \in L$, define

$$\varphi \precsim_{\Sigma}^{\mathrm{PL}} \psi \Leftrightarrow \Sigma, \varphi \models_{\mathrm{PL}} \psi.$$

The derived relation $\precsim_{\Sigma}^{\mathrm{PL}}$ thus ranks sentences of L by their deductive strength relative to the theory Σ. As usual, define $\varphi \sim_{\Sigma}^{\mathrm{PL}} \psi$ as both $\varphi \precsim_{\Sigma}^{\mathrm{PL}} \psi$ and $\psi \precsim_{\Sigma}^{\mathrm{PL}} \varphi$, and define $\varphi \prec_{\Sigma}^{\mathrm{PL}} \psi$ as $\varphi \precsim_{\Sigma}^{\mathrm{PL}} \psi$ and $\psi \not\precsim_{\Sigma}^{\mathrm{PL}} \varphi$.

Definition 4.1 For a propositional language L, $\Sigma \subseteq L$, $\varphi \in L$, define

$$[\![\varphi]\!]_{\Sigma} = \{v \in \mathrm{Models}(L) \mid v(\varphi) = T \text{ and for any } \psi \in \Sigma, v(\psi) = T\}.$$

$[\![\varphi]\!]_{\Sigma}$ is thus the collection of Σ-models which also make φ true.

Proposition 4.1

For a propositional language L, consistent $\Sigma \subseteq L$, and $\varphi, \psi \in L$,

$$\varphi \precsim_{\Sigma}^{PL} \psi \Leftrightarrow [\![\varphi]\!]_{\Sigma} \subseteq [\![\psi]\!]_{\Sigma}.$$

Presented in terms of models, the defined $\precsim_{\Sigma}^{\mathrm{PL}}$ relation corresponds to the subset relation over sets of Σ-models. It follows immediately that $\precsim_{\Sigma}^{\mathrm{PL}}$ has the properties of the subset-relation over this space:

Proposition 4.2

For a propositional language L, consistent $\Sigma \subseteq L$, and $\varphi, \psi \in L$,

(i) Non-triviality
$\bot \prec_{\Sigma}^{PL} \top$.

(ii) Transitivity
For any $\varphi_1, \varphi_2, \varphi_3 \in L$, if $\varphi_1 \precsim_{\Sigma}^{PL} \varphi_2$ and $\varphi_2 \precsim_{\Sigma}^{PL} \varphi_3$, then $\varphi_1 \precsim_{\Sigma}^{PL} \varphi_3$.

(iii) Reflexivity
For any $\varphi \in L$, $\varphi \precsim_{\Sigma}^{PL} \varphi$.

(iv) Boundedness
For any $\varphi \in L$ such that $\varphi \nsim_{\Sigma}^{PL} \bot$ and $\varphi \nsim_{\Sigma} \top$, $\bot \prec_{\Sigma}^{PL} \varphi \prec_{\Sigma}^{PL} \top$.

(v) Monotonicity
If $\varphi \wedge \gamma \sim_{\Sigma}^{PL} \psi \wedge \gamma \sim_{\Sigma}^{PL} \bot$, then $\varphi \precsim_{\Sigma}^{PL} \psi$ if and only if $\varphi \vee \gamma \precsim_{\Sigma}^{PL} \psi \vee \gamma$.

Given any propositional language L and any consistent L-theory Σ, the propositional consequence relation induces a comparative ranking over the sentences of L. There thus exists a large stock of widely-accepted confirmation relations of varying degree. While a specific value in an absolute scale may be difficult to identify, it is nevertheless obvious that, with no evidence, $P \vee Q$ is better confirmed than P, $P \wedge Q$ is less confirmed than P, and so forth. Ramsey's skepticism is thus unwarranted. Propositional logic itself provides an example—albeit a weak one—of the kind of theory sought.

First-Order Logic as Confirmation Theory

First-order logic can likewise be viewed as a minimalist confirmation theory. Let L be a first-order language and let $\models_{\text{FOL}}$ denote the first-order consequence relation. For consistent $\Sigma \subseteq L$ and any $\varphi, \psi \in L$, define

$$\varphi \precsim_{\Sigma}^{\text{FOL}} \psi \Leftrightarrow \Sigma, \varphi \models_{\text{FOL}} \psi.$$

The derived relation $\precsim_{\Sigma}^{\text{FOL}}$ again ranks sentences of L by their deductive strength relative to the theory Σ. $\varphi \sim_{\Sigma}^{\text{FOL}} \psi$ will again be taken as an abbreviation for both $\varphi \precsim_{\Sigma}^{\text{FOL}} \psi$ and $\psi \precsim_{\Sigma}^{\text{FOL}} \varphi$ while $\varphi \prec_{\Sigma}^{\text{FOL}} \psi$ is shorthand for $\varphi \precsim_{\Sigma}^{\text{FOL}} \psi$ and $\psi \not\precsim_{\Sigma}^{\text{FOL}} \varphi$. The sets of models characterization provided in the propositional case is unavailable in the general first-order case because the collection of models which make Σ true need not form a set. All of the characteristics are nevertheless retained:

Proposition 4.3

For a first-order language L, consistent $\Sigma \subseteq L$, and $\varphi, \psi \in L$,

(i) Non-triviality
$\bot \prec^{FOL}_{\Sigma} \top$.

(ii) Transitivity
For any $\varphi_1, \varphi_2, \varphi_3 \in L$, if $\varphi_1 \precsim^{FOL}_{\Sigma} \varphi_2$ and $\varphi_2 \precsim^{FOL}_{\Sigma} \varphi_3$, then $\varphi_1 \precsim^{FOL}_{\Sigma} \varphi_3$.

(iii) Reflexivity
For any $\varphi \in L$, $\varphi \precsim^{FOL}_{\Sigma} \varphi$.

(iv) Boundedness
For any $\varphi \in L$ such that $\varphi \nsim^{FOL}_{\Sigma} \bot$ and $\varphi \nsim^{FOL}_{\Sigma} \top$, $\bot \prec^{FOL}_{\Sigma} \varphi \prec^{FOL}_{\Sigma} \top$.

(v) Monotonicity
If $\varphi \wedge \gamma \sim^{FOL}_{\Sigma} \psi \wedge \gamma \sim^{FOL}_{\Sigma} \bot$, then $\varphi \precsim^{FOL}_{\Sigma} \psi$ if and only if $\varphi \vee \gamma \precsim^{FOL}_{\Sigma} \psi \vee \gamma$.

Given any first-order language L and any consistent L-theory Σ, the first-order consequence relation thus induces a comparative ranking over L which satisfies all the axioms of intuitive probability save comparability. Despite Ramsey's skepticism, philosophers have been in possession of an objective and normative theory of confirmation for decades. The only question is whether confirmation can be rigorously strengthened beyond classical logic.

4.3 The Equipossibility Account

The classical conception of probability suggests a simple method for extending confirmation beyond logic. For an extended overview of this period, Daston (1995) and Hacking (2006) provide excellent and contrasting accounts. The most iconic proponent of classical probability was the French mathematician Pierre-Simon Laplace who, following Gottfried Leibniz and Jakob Bernoulli, explicitly defined probability as degree of possibility[2] (Hacking 1971). In Laplace's ([1814] 1951) conception,

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability,

[2] As Hacking (1971, 2006) argues at length, there is good reason to worry over the meaning(s) of 'possibility' in these works. Authors in this period tend to switch between a physical interpretation (ease of bringing about) and an epistemic interpretation (consistency with the evidence).

which is, thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible. (6)

Laplace offers here a two-step account of probability:

(1) If evidence is symmetrically distributed among cases, then these cases are equally possible.

(2) The probability of a proposition P is the number of equally possible cases which make P true divided by the total number of equally possible cases.

Probability is thus derived from equal possibility or 'equipossibility' between cases.

This equipossibility account is the basis both for Laplace's use of (POI) and his identification of probabilities with numerical values ranging between zero and one. (POI) asserts that symmetric evidence between alternatives entails equal probability. Given Laplace's commitment to (1), symmetric evidence between alternatives requires either that these alternatives are themselves Laplace's equally possible cases or that these alternatives have Laplace's equally possible cases distributed uniformly among them. In either situation, Laplace's definition of probability—(2) above—entails that each alternative is equally probable. (POI) is thus a straightforward consequence of pairing (1) and (2). Similarly, if the probability of a proposition is the ratio of equal possibilities favorable to a proposition to the total number of equal possibilities, then at minimum there are no favorable possibilities and at maximum all possibilities are favorable. The minimal and maximum probabilities are thus zero and one respectively with intermediate sets appearing between these two bounds. The equipossibility account is thus simultaneously the basis for the most widely accepted and the most controversial aspects of classical probability.

Two centuries before the widely acknowledged foundations of confirmation, Laplace not only alleges the existence of a degreed relation which is evidence-relative, objective, and normative but also provides an extended account of this relation. Further, Laplace regularly applies this account to problems characteristic of confirmation. The most infamous instance here is the likelihood that the sun will rise tomorrow given only that it has done so in the past, already a default example in discussions of induction by Laplace's time (Zabell 1989). Provided both Laplace's explicit commitments and his application of the account, there is good reason to regard the equipossibility account—and by extension the mathematical theory of probability—as an early analysis of confirmation. Laplace is not investigating a particular axiomatic system dubbed 'probability' but rather formalizing the same intuitions that would later drive Keynes, Jeffreys, and Carnap. Popular history has gotten the relation between confirmation and probability backwards.

Early theory of confirmation or not, the equipossibility account is widely held to be circular (von Mises 1957; Reichenbach 1971; Hacking 2006). The worry is that equally possible cases are equally probable cases not because the equipossibility account is correct but rather because the notion of equipossibility is a flagrant repackaging of equal probability. This account then presupposes an antecedent notion of probability, and so no ground has been gained. This circularity charge is accepted even among contemporary commentators who are favorably disposed towards the equipossibility account. Rosenkrantz (1981), for example, accepts that Laplace means equiprobable by 'equipossible' but maintains—despite Laplace's explicit claim to the contrary—that (2) is not offered as a definition. Defenses of this kind have rightfully proven unpersuasive, and so the equipossibility account has been relegated to a historical note.

4.3.1 Finite, Atomic Confirmation for Propositional Languages

Despite this, a non-circular variant of the equipossibility account is readily available. Replacing Laplace's talk of cases with an exhaustive collection of atomic[3] possibilities and Laplace's appeal to symmetrically distributed evidence with taking all atomic possibilities consistent with the evidence, (1) claims that all and only the atomic possibilities consistent with our evidence are equally possible.[4] (2) need only be modified to talk explicitly about degrees of confirmation over collections of these atomic possibilities. Restricted to propositional languages, this gives:

> FINITE, ATOMIC CONFIRMATION FOR PROPOSITIONAL LANGUAGES
>
> Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set of evidence.
>
> (1') If $\Omega_{\mathcal{L}}$ is an exhaustive collection of atomic possibilities, then every possibility in the subcollection $\Omega_{\mathcal{L},\Sigma}$ is an equal possibility relative to $\Omega_{\mathcal{L}}$ and Σ.
>
> (2') For a finite, non-empty set of equal possibilities, the degree of confirmation for a sentence φ is the number of equal possibilities which make φ true divided by the total number of equal possibilities.

Given a finite set of atomic possibilities and a set of evidence from some propositional language L, this interpretation of the equipossibility account defines a degree of confirmation for every sentence of L. The talk of equal possibilities here is evocative but inessential; confirmation is ultimately reduced to the collection of atomic possibilities that are consistent with the evidence.

The atomic confirmation account described by (1') and (2') nevertheless coheres with characteristic applications of the equipossibility account. Suppose, for example, that we are faced with exactly two logically independent propositions P and Q, most famously 'the coin lands heads on the first toss' and 'the coin lands heads on the second toss'. Exactly four propositional models are available here:

[3]An *atomic* possibility is a complete or indivisible way the world could be. Note that this is consistent with both the existence and non-existence of a set of atomic elements within a given possibility.

[4]This modification was anticipated by Keynes (1921) in his analysis of the principle of indifference.

	P	Q
v_1 :	T	T
v_2 :	T	F
v_3 :	F	T
v_4 :	F	F

Each of these gives a way the world could be with respect to P and Q.

Suppose further that we regard each of these propositional models as atomic, as describing an irreducible way the world could be. With the current example, this is clearly false; the world contains more than a single coin landing either heads or tails. The current example is nevertheless instructive, and failures of atomicity need not always occur. Still further, there is significant intuitive appeal to *regarding* a set of possibilities as atomic when we are faced with a question which doesn't require further detail. Even if we eventually come to recognize this as an error, it is an intuitively appealing error that helps makes sense of both early commentaries and pretheoretic intuitions.

Lastly, suppose that we are supplied no evidence as to the outcome of either the first coin toss or the second. Each of the atomic propositional models v_1-v_4 is thus by (1') to be accounted equally possible. Following (2') and defining degrees of confirmation as the ratio of favorable equal possibilities to all equal possibilities:

Sentence	Atomic Models	Degree of Confirmation
$\bot$	$\emptyset$	0
$P \wedge \neg Q$	$\{v_2\}$	$\frac{1}{4}$
P	$\{v_1, v_2\}$	$\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
$P \vee Q$	$\{v_1, v_2, v_3\}$	$\frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$
$P \vee \neg P$	$\{v_1, v_2, v_3, v_4\}$	$\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$

The cumulative effect is a probabilistic ranking of propositional sentences which accords with the expected frequencies for a fair coin.

Given a propositional language L, a finite model space $\Omega_{\mathcal{L}}$ for L, and a consistent (with respect to $\Omega_{\mathcal{L}}$) set of evidence $\Sigma \subseteq L$, (1') and (2') supply a unique degree of confirmation for every sentence in L.

Definition 4.2 For L a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ consistent (with respect to $\Omega_{\mathcal{L}}$), the *atomic confirmation ranking* is

$$\mu_{\Sigma}^{\mathcal{L}}(\varphi) = \frac{|[\![\varphi]\!]_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}$$

where $\Omega_{\mathcal{L},\Sigma}$ is the set of models from $\Omega_{\mathcal{L}}$ which make Σ true while $[\![\varphi]\!]_{\mathcal{L},\Sigma}$ is the set of models from $\Omega_{\mathcal{L}}$ which make both Σ and φ true.

Definition 4.3 For L a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ consistent (with respect to $\Omega_{\mathcal{L}}$), the *comparative atomic confirmation ranking* is defined by

$$\varphi \precsim^{\mathcal{L}}_{\Sigma} \psi \Leftrightarrow \mu^{\mathcal{L}}_{\Sigma}(\varphi) \leq \mu^{\mathcal{L}}_{\Sigma}(\psi).$$

As in Chapter 2, a formal language L with Boolean connectives together with a collection of models $\Omega_{\mathcal{L}}$ defined over L straightforwardly gives rise to the algebra $\mathcal{F}_{\mathcal{L}} = \{[\![\varphi]\!]_{\Omega_{\mathcal{L}}} : \varphi \in L\}$. A formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a consistent (with respect to $\Omega_{\mathcal{L}}$) set $\Sigma \subseteq L$ similarly gives rise to an algebra $\mathcal{F}_{\mathcal{L},\Sigma} = \{[\![\varphi]\!]_{\Omega_{\mathcal{L},\Sigma}} : \varphi \in L\}$. If we wish to abstract from the expressive limitations of L, we can generalize these to $\mathcal{F}^{+}_{\mathcal{L}} = \mathcal{P}(\Omega_{\mathcal{L}})$ and $\mathcal{F}^{+}_{\mathcal{L},\Sigma} = \mathcal{P}(\Omega_{\mathcal{L},\Sigma})$ respectively. Both $\mu^{\mathcal{L}}_{\Sigma}$ and $\precsim^{\mathcal{L}}_{\Sigma}$ may be unambiguously extended to $\mathcal{F}^{+}_{\mathcal{L}}$ by setting

$$\mu^{\mathcal{L}}_{\Sigma}(E) = \frac{|E \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}$$

for any $E \in \mathcal{F}^{+}_{\mathcal{L}}$.

Just like Laplace's presentation of the equipossibility account, the atomic confirmation account described by (1') and (2') is closely related to the mathematical theory of probability.

Proposition 4.4

Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then, $\mu^{\mathcal{L}}_{\Sigma}$ satisfies Kolmogorov's probability axioms over $\mathcal{F}^{+}_{\mathcal{L}}$.

Upon learning some φ consistent with $\Omega_{\mathcal{L},\Sigma}$, the account also endorses conditionalization.

Proposition 4.5

Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set, and φ a sentence consistent with $\Omega_{\mathcal{L},\Sigma}$. Then, for any ψ,

$$\mu^{\mathcal{L}}_{\Sigma}(\psi|\varphi) = \mu^{\mathcal{L}}_{\Sigma\cup\{\varphi\}}(\psi).$$

Atomic confirmation thus leads us directly to the hallmarks of contemporary objective Bayesianism. If (1') and (2') are a plausible formalization of (1) and (2), this is no accident. Much of the perennial appeal of objective Bayesian reasoning is underwritten by implicit appeals to (1') and (2').

History aside, (1') and (2') provide a comparative confirmation ranking which strictly extends the comparative confirmation relation induced by propositional logic.

Proposition 4.6

Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then,

$$\varphi \precsim_{\Sigma}^{PL} \psi \Rightarrow \varphi \precsim_{\Sigma}^{\mathcal{L}} \psi$$

$$\varphi \precsim_{\Sigma}^{\mathcal{L}} \psi \nRightarrow \varphi \precsim_{\Sigma}^{PL} \psi.$$

The intuition behind this extension is interchangeability. Given a set of atomic models, the subsequent restriction to models consistent with the evidence supplies a set of possibilities which are on par in all relevant dimensions. Every such model corresponds to precisely one way things could be, and no such model is in a better position vis-á-vis the evidence. All that then distinguishes more confirmed sentences from less is the corresponding number of atomic models satisfying Σ.

Contrary to the modern consensus, the charge of circularity against the equipossibility account of probability can be coherently evaded. Indeed, while Laplace lacked the logical framework underlying (1') and (2'), the retreat to atomic possibility itself is exceedingly natural, particularly in the context of games of chance.[5] Beyond this historical point, atomic confirmation for propositional languages provides a natural extension of confirmation theory beyond propositional logic. Given a set of atomic models, any model from this set which is consistent with our evidence is as good as any other.

4.4 D'Alembert's Riddle

The atomic confirmation account sketched in the previous section depends on either being supplied or identifying a set of atomic models. While in some cases we might be explicitly supplied such a set, this seems the exception rather than the rule. Is it possible, then, to identify a set of atomic models? The first suggestion that it is not comes from a "riddle" posed by the 18th century French mathematician Jean-Baptiste d'Alembert. While contemporary criticism of the equipossibility account centers on its perceived circularity, d'Alembert focused instead on Laplace's talk of cases, offering a simple two-player coin game to illustrate his misgivings.

In d'Alembert's proposed game, the first player wins just in case the coin lands heads within two tosses (Todhunter 1865). If the coin fails to land heads within two tosses, the second player wins instead. When asked for the space of equipossibilities in this game,

[5] Is a retreat to atomic possibility what Laplace really intended all along? No, not in any explicit sense. This is supported by the injunction that we need only reduce "all the events of the same kind" to equal possibilities in the quoted passage. More generally, there is a salient intuitive notion of physical possibility (or ease of accomplishment) which appears to underwrite some of Laplace's intuitions (Hacking 2006). For an extended development of this idea, see Myrvold (2019).

most commentators—Laplace among them—provide a space which fixes outcomes for two tosses of the coin:

$$\text{HH} \mid \text{HT} \mid \text{TH} \mid \text{TT}$$

Applying (2'), the proposition 'at least one coin toss lands heads' or $HH \vee HT \vee TH$ receives degree of confirmation $\frac{3}{4}$. This solution is challenged by d'Alembert who notes that a second coin toss needs occur in this game only if the first toss is tails. He therefore contends that the space of equipossibilities is instead:

$$\text{H} \mid \text{TH} \mid \text{TT}$$

It then follows that the degree of confirmation for 'at least one coin toss lands heads' or $H \vee TH$ is $\frac{2}{3}$.

Modern commentaries tend to treat d'Alembert's proposed solution as a quaint misstep. As many have been quick to point out, actually playing d'Alembert's game a large number of times gives frequencies which accord with the standard response of $\frac{3}{4}$ and not d'Alembert's value of $\frac{2}{3}$. A small amount of probability theory likewise shows that a fair coin lands heads at least once with probability $\frac{3}{4}$. Despite their popularity, both of these responses to the riddle are inadequate. The coin at issue in the puzzle is not indicated to be fair or indeed biased in any particular manner; it is likewise not required that the coin used is similar to any coin with which we are acquainted.[6] Both probability calculations which suppose a fair coin and the outcomes of experiments are thus irrelevant to d'Alembert's game. A different bias and a different structure for common coins will have both probability theory and experimental results favoring d'Alembert's solution instead.

More generally, d'Alembert's riddle—like many puzzles surrounding confirmation[7]—suffers from the ready availability of evidence beyond that explicitly provided by the riddle itself. Both reactions above stem from natural though illicit attempts to supplement the evidence available in the riddle with our own experiences. The tossing of coins is something each of us is intimately familiar with, and so there is intense pressure to utilize this additional information in evaluating the likelihood that at least one coin toss lands heads. Students of probability theory have an even greater difficulty; not only must they contend with their own experiences tossing coins but also training to interpret 'coin toss' as code for a very particular kind of probabilistic event. Nevertheless, understanding d'Alembert's riddle requires setting aside this illicit background evidence. All that is intended in the riddle is that we are faced with a process which is certain, once invoked, to resolve in one of exactly two ways.

[6]Whether or not d'Alembert intentionally omitted these assumptions is unclear; in all likelihood, d'Alembert simply lacked the sharp distinctions between probabilities, frequencies, chances, and degrees of confirmation typical of modern work. Intentional or not, this omission is crucial to the viability of d'Alembert's solution.

[7]Most famously, *The Paradox of the Ravens* presented in Hempel (1945).

Rather than a simple error on d'Alembert's part, the conflicting solutions to the riddle stem from two incompatible interpretations of how the world will be by the game's conclusion. The standard response imagines that the coin is tossed twice whereupon a victor is crowned. In contrast, d'Alembert's proposal makes use of the fact that if the coin initially lands heads, a second toss would be irrelevant to the outcome of the game. With this in mind, the game may instead be played by making a first toss and then a second only if the game is not already settled. Nothing in d'Alembert's description stipulates one version or the other; the choice here is apparently arbitrary.

Accepting d'Alembert's proposed solution as legitimate, the classical equipossibility account no longer supplies an unambiguous probability for player 1's victory. This failure ultimately derives from indeterminacy in the notion of 'case' or 'possibility' invoked by the account. D'Alembert's crucial contribution here is an intuitively acceptable yet non-standard space of possibilities. Unlike the classical equipossibility account, however, (1') and (2') explicitly derive degrees of confirmation from a set of atomic possibilities. In this context, d'Alembert's riddle fails to motivate any indeterminacy. If the four possibilities laid out in the standard solution are the atomic possibilities, then the standard solution is correct. Similarly, if the three possibilities laid out in d'Alembert's solution are the atomic possibilities, then d'Alembert's solution is correct. D'Alembert's riddle nevertheless casts doubt on the possibility of identifying atomic possibilities. This point does not undermine the atomic confirmation account as a whole because atomic possibilities are at least sometimes explicitly stipulated. It does, however, greatly restrict the account's applicability if all choices outside such cases are subjective. D'Alembert's riddle thus represents a substantial threat to the atomic confirmation account presented in the previous section.

It is tempting to pin the conflict between the standard solution and d'Alembert's on a simple, one-shot ambiguity between the two proposed solutions. Natural language sentences often admit of multiple logically distinct interpretations; the sentence 'everyone loves someone' may, for instance, be understood as asserting either that there is some single individual loved by everyone ($\exists x \forall y L(y, x)$) or that for any person, there is someone they love ($\forall y \exists x L(y, x)$). Similarly, the description of d'Alembert's riddle may be ambiguous between two well-defined games: one where two tosses of the coin always occur and one where the second toss occurs only if necessary. On this treatment, there would be no true conflict between the two proposed solutions and no robust challenge to either equipossibility account. Each solution targets one of the two viable interpretations for the provided description.

Unfortunately, the phenomenon underlying d'Alembert's riddle runs far deeper than only two incompatible interpretations. D'Alembert's description of his game is in fact consistent with an infinite number of interpretations. For a particularly striking example, note that nothing in d'Alembert's description of the game rules out the first player being required to clap if the coin lands heads on the first toss. Incorporating the possibility of this clapping (C) into the standard analysis now gives six possibilities:

$$\text{HHC} \mid \text{HTC} \mid \text{THC} \mid \text{TTC} \mid \text{TH}\neg\text{C} \mid \text{TT}\neg\text{C}$$

Of these six, four possibilities have the coin landing heads at least once. If these are our atomic possibilities, then the degree of confirmation associated with 'at least one coin toss lands heads' is $\frac{2}{3}$. If we instead follow d'Alembert and minimize the number of coin flips, the proper degree of confirmation for the game with clapping is $\frac{3}{5}$. Regardless of our position on coin flips, the introduction of clapping produces a shift in how likely victory is for the first player.

Even amongst those friendly to d'Alembert's original challenge, there is great pressure to declare this new example defective. While the original riddle was induced by a part of the game itself, the first player's clapping is a clearly irrelevant addition. Accordingly, the introduction of clapping into the possibilities considered is illegitimate, and the resulting degrees of confirmation need not be taken seriously. Despite the intuitive appeal of this response, it repeats the error flagged at the outset of this section. Given only the evidence provided, the relevance of clapping to the outcome of the coin toss is entirely undetermined. Only by helping ourselves to additional, illicit evidence (e.g., observed non-interaction between quarters and clapping) does the outright dismissal of the new clapping proposition become attractive. Given only d'Alembert's specification of the game, the first player's clapping might well be of immense importance.

Recognizing this, there is a clear parallel between the standard assumption that the coin must be tossed twice and the assumption that the first player must clap if the first toss lands heads. Neither of these assumptions are necessary for the completion of the game d'Alembert described. Similarly, neither assumption is explicitly endorsed nor explicitly rejected by d'Alembert's description. Indeed, the only robust disanalogy between these two is the feeling of naturalness and irrelevance respectively delivered by an illicit appeal to our own experiences. Neither is properly relevant to the riddle, and so requiring clapping for a heads result on the first toss is on par with requiring a second coin flip. If the standard solution is legitimate, so too are both clapping solutions.[8]

Examples of this kind are easily multiplied. In fact, 'at least one coin lands heads' can be brought arbitrarily close to both zero and one with different choices for atomic possibilities. D'Alembert has identified not a one-shot ambiguity but a general phenomenon that threatens to undermine atomic confirmation in any situation which is not antecedently equipped with a set of atomic possibilities.

4.4.1 Propositional Formalization of D'Alembert's Riddle

The atomic confirmation account presented in the previous section makes extensive use of the framework developed for propositional logic, and the standard solution to d'Alembert's riddle is particularly compelling when presented in these terms. Formulating and motivating d'Alembert's solution in this framework not only helps to clarify the riddle itself but also the riddle's relationship to propositional logic.

[8]The outlier here is actually d'Alembert's solution rather than the standard proposal. Given the specification of the game, every possibility d'Alembert recognizes is necessary. The same cannot be said for any of other solutions proposed in this section.

Formally, the discrepancy between the standard solution and d'Alembert's alternative appears with the initial translation into a formal language:

Standard Solution
H_1: The coin lands heads on the first toss.
T_1: The coin lands tails on the first toss.
H_2: The coin lands heads on the second toss.
T_2: The coin lands tails on the second toss.

D'Alembert's Solution
H_1: The coin lands heads on the first toss.
T_1: The coin lands tails on the first toss.
H_2^D: There is a second toss, and the coin lands heads on it.
T_2^D: There is a second toss, and the coin lands tails on it.

Whether or not H_2 and H_2^D or T_2 and T_2^D are properly equivalent, the two solutions definitively differ on the meaning postulates adopted for these propositional atoms.

Standard Solution	D'Alembert's Solution
$H_1 \leftrightarrow \neg T_1$	$H_1 \leftrightarrow \neg T_1$
$H_2 \leftrightarrow \neg T_2$	$H_2^D \rightarrow \neg T_2^D$
	$T_2^D \rightarrow \neg H_2^D$
	$H_1 \rightarrow \neg T_2^D \wedge \neg H_2^D$

On the standard solution, not landing heads on the second toss is taken to imply landing tails and vice-versa for landing tails on the second toss. On d'Alembert's solution meanwhile neither implication holds. Indeed, landing heads on the first toss implies that no second toss occurs. While the former scheme is simpler, neither is clearly wrong. The established discrepancy in atomic models is readily apparent:

	H_1	T_1	H_2	T_2
v_1 :	T	F	T	F
v_2 :	T	F	F	T
v_3 :	F	T	T	F
v_4 :	F	T	F	T

	H_1	T_1	H_2^D	T_2^D
v_1^D :	T	F	F	F
v_2^D :	F	T	T	F
v_3^D :	F	T	F	T

The critical behavior here thus arises not only with a choice of atomic models but also with the selection of propositional atoms and meaning postulates.

In propositional logic, propositional atoms are characterized by (i) their restriction to exactly one of two truth values—true and false—and (ii) their lack of further structure. The contributions of the first are explicit in the accompanying formal semantics while the implications of (ii) have often—as Burgess (2003) argues at length—been misinterpreted by even sophisticated commentators. Despite the unstructured nature of propositional atoms, these symbols need not correspond to either contingent or atomic sentences. Rather, propositional atoms correspond to sentences of arbitrary complexity (propositional or otherwise). It is thus entirely legitimate to associate propositional atoms with complex sentences as in H_2^D and T_2^D. As a direct result, particular propositional atoms may be assessed as true in a model while the associated sentence is necessarily false or assessed as false while the associated sentence is necessarily true. Propositional logic thus exhibits a fundamental blindness towards the actual sentences associated with propositional atoms.

The blindness induced by (ii) is simultaneously a critical advantage and a severe limitation of propositional logic. In its favor, this blindness allows the formalism to advance without consideration of semantic and syntactic complexities beyond those explicitly introduced. It is this disconnect that makes the simplicity of the propositional account possible. At the same time, many of these suppressed complexities are also the basis for legitimate inferences. Propositional logic therefore suffers from an immediate yet necessary shortfall as a result of (ii); some truly legitimate inferences have been sacrificed in order for the account as a whole to proceed.

Thankfully, this shortfall can be partially repaired. While propositional atoms lack further structure, additional truth-functional information may be directly encoded in the propositional formalism by imposing outside restrictions—meaning postulates—on propositional atoms. This is precisely what occurs above, for instance, when ‘the coin lands heads on the first toss’ and ‘the coin lands tails on the first toss’ are associated with the propositional atoms H_1 and T_1 respectively and then immediately supplemented by a claim relating these. In isolation, propositional logic recognizes four possibilities for the pair of atoms:

H_1	T_1
T	T
T	F
F	T
F	F

Introducing the meaning postulate $H_1 \leftrightarrow \neg T_1$, however, rules out the first and last. In this manner, some of the information within the original sentences is reintroduced despite (ii).

The atomic confirmation account upends this entire approach; the desired models are now the truly atomic ones, and any simplification compromises the resulting degrees of confirmation. D’Alembert’s contribution is only to dramatize the difficulties inherent in this demand for atomicity. Both the standard solution and d’Alembert’s alternative accept the syntactic description of d’Alembert’s game but interpret this description differently. Nothing in the syntax of the sentences themselves marks out one interpretation or the other as correct, and so it seems that both must be accounted legitimate. As the formalization above illustrates, the contrast between these two interpretations can be viewed either as a disagreement over which set of atomic models to adopt or as a disagreement over the appropriate choice of formal language and meaning postulates. On this latter formulation, proponents of each solution are naturally regarded as speaking syntactically identical but semantically distinct languages.

4.4.2 D'Alembert's Riddle and Interpreted Propositional Languages

Formulating d'Alembert's riddle as a disagreement in language suggests a simple characterization of when changes in the atomic confirmation ranking are and are not to be expected. For simplicity, we restrict our attention in this subsection to model spaces $\Omega_{\mathcal{L}}$ that are composed entirely of L-models.

Definition 4.4 An interpreted propositional language $\mathcal{L}$ is a pair of a propositional language L and a set of L-models $\Omega_{\mathcal{L}}$.

The standard solution and d'Alembert's alternative thus differ in the interpreted language they endorse. In many cases, atomic model spaces may be generated by specifying a consistent set of meaning postulates $\Sigma_M \subseteq L$ and then taking all L-models consistent with this set. A formal language L, exhaustive collection of L-models, and set of meaning postulates thus suffice to generate an interpreted language though the reverse is not always true.

The simplest relationship between two distinct interpreted languages is extension:

Definition 4.5 An interpreted propositional language $\mathcal{L}^+$ *extends* an interpreted language $\mathcal{L}$ if and only if L^+ extends L and for any $v^+ \in \Omega_{\mathcal{L}^+}$, $v^+|_L \in \Omega_{\mathcal{L}}$.

While the language difference underpinning d'Alembert's riddle is not an instance of extension, any difference between interpreted languages can be viewed as a succession of extensions from a common base language. Understanding the behavior of atomic confirmation rankings under interpreted language extensions can thus supply a general understanding of when shifts in the atomic confirmation ranking can be expected.

The atomic confirmation ranking proves resilient to interpreted language extensions in only two general cases. The first of these cases is the extension of an interpreted language by a symbol which is already—barring the expressive weaknesses of the language—representable.

Definition 4.6 Given an interpreted propositional language $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$, an extension $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+} \rangle$ of $\mathcal{L}$ is *completely restricted* if and only if for every new propositional letter P in L^+, there exists $R_P \subseteq \Omega_{\mathcal{L}}$ such that

$$[\![P]\!]_{\Omega_{\mathcal{L}^+}} = \{v^+ \in \Omega_{\mathcal{L}^+} \ : \ \text{for some } v \in R_P, v^+|_L = v\}.$$

The propositions introduced by a completely restricted extension are equivalent to some already existing subset of $\Omega_{\mathcal{L}}$. Every atomic model for the extended language $\mathcal{L}^+$ then corresponds to exactly one atomic model from the original language $\mathcal{L}$.

Proposition 4.7

For any interpreted propositional languages $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$ and $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+} \rangle$, if $\mathcal{L}^+$ is a completely restricted extension of $\mathcal{L}$, then there exists a bijection $\delta : \Omega_{\mathcal{L}} \to \Omega_{\mathcal{L}^+}$ which preserves the truth of all L-sentences.

The atomic confirmation ranking is then fixed under this extension whenever it is defined, viz. when the set of atomic models consistent with the evidence Σ is finite.

Proposition 4.8

Let interpreted propositional languages $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$ and $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+} \rangle$ as well as a consistent (with respect to $\Omega_{\mathcal{L}}$) theory Σ be given. If $\mathcal{L}^+$ is a completely restricted extension of $\mathcal{L}$ and $\Omega_{\mathcal{L}^+,\Sigma}$ is finite and non-empty, then for any set $A \in \mathcal{F}^+_{\mathcal{L}}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{v^+ \in \Omega_{\mathcal{L}^+} \;:\; v^+|_L \in A\}$.

In the case where $A = [\![\varphi]\!]_{\Omega_{\mathcal{L},\Sigma}}$, we thus have that $\mu^{\mathcal{L}}_{\Sigma}(\varphi) = \mu^{\mathcal{L}^+}_{\Sigma}(\varphi)$. Completely restricted extensions of interpreted propositional languages leave the atomic confirmation ranking fixed.

The second fixed case for the atomic confirmation ranking is unrestricted extensions.

Definition 4.7 Given an interpreted propositional language $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$, an extension $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+} \rangle$ of $\mathcal{L}$ is *unrestricted* if and only if

$$\Omega_{\mathcal{L}^+} = \{L^+ \text{ valuations } v^+ \;:\; v^+|_L \in \Omega_{\mathcal{L}}\}.$$

Unrestricted extensions introduce propositional atoms without any accompanying meaning postulates or, equivalently, with no restrictions on the truth values assigned.

Proposition 4.9

Let $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$ be an interpreted propositional language and $\Sigma \subseteq L$ a theory such that $\Omega_{\mathcal{L},\Sigma}$ is finite and nonempty. If $\mathcal{L}^+ = \langle L, \Omega_{\mathcal{L}^+} \rangle$ is a finite, unrestricted extension of $\mathcal{L}$, then for any $A \in \mathcal{F}^+_{\mathcal{L},\Sigma}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{v^+ \in \Omega_{\mathcal{L}^+} \;:\; v^+|_L \in A\}$.

In the case where $A = [\![\varphi]\!]_{\Omega_{\mathcal{L},\Sigma}}$, we thus have that $\mu^{\mathcal{L}}_{\Sigma}(\varphi) = \mu^{\mathcal{L}^+}_{\Sigma}(\varphi)$. Unrestricted extensions of interpreted propositional languages also leave the atomic confirmation ranking fixed.

Not all extensions of interpreted languages which fix the atomic confirmation ranking fall into the two general cases presented. These further instances, however, have an air of happenstance, a series of additions which would individually shift the ranking but when combined happen to balance. Accordingly, further instances of interpreted language sensitivity can typically be generated simply by avoiding completely restricted and unrestricted extensions. The mechanism underlying d'Alembert's riddle is then a general feature of the atomic confirmation ranking for propositional languages. Without a specified set of atomic possibilities, inconsistent interpreted propositional languages can be readily generated by adding propositional atoms whose truth values are only partially restricted.

4.5 Finite Confirmation for First-Order Languages

The cumulative effect of the previous section is a robust challenge to any application of the atomic confirmation account for propositional languages when a set of atomic possibilities is not explicitly specified as part of our evidence. Further, this was cast as a severe limitation of the theory. The current section sketches a two-part response to the issues raised by d'Alembert's riddle. First, first-order logic supplies the resources for an exhaustive characterization of atomic possibilities. Transitioning to a first-order account of atomic confirmation thus allows for the identification of a collection of atomic possibilities given any set of evidence. Second, confirmation relations ought to exhibit a general sensitivity to changes in the collection of atomic models under consideration. Appearances aside, the discrepancy between the standard solution and d'Alembert's proposal reveals a radical disagreement about the nature of the world.

4.5.1 First-Order Logic

If we are speaking entirely meaningfully, our use of a name picks out a unique object, our use of a relation-name picks out a unique relation, and so forth. In so far as these criteria fail, our words likewise fail to have a single, well-defined meaning; we speak to some degree vaguely and ambiguously. The large part of human discourse falls into this latter category; complete precision is a high standard, and it is typically good enough to just narrow down the available extensions of our symbols. It is of little practical consequence after all whether a particular rock formation on Pluto qualifies as a chair, and accordingly little attention is paid to its position vis-á-vis the extension of 'chair'. So far as the important objects around us are sorted appropriately into chairs and non-chairs, 'chair' is meaningful enough.

First-order logic leverages this situation into an account of logical consequence. If unambiguous or ideal use of a language requires appropriate extensions for each component of the language, then the structure of these extensions induces restrictions on what follows from the truth of particular sentences. The parallel with propositional logic is striking here, and the progression from this point appears to parallel the propositional case. Supplied with a formal language, a collection of models exhausting all possible extensions can be constructed. A sentence φ can then be defined as a logical consequence of a sentence ψ just in case every model in the exhaustive collection which makes φ true also makes ψ true.

The only difference between propositional and first-order logic appears, then, to be the shift from extensions of sentences to extensions of parts of sentences.

As Etchemendy (1990) observes, the foundations of first-order logic are not quite so straightforward as this. Since the extensions our symbols could have are naturally constrained by which objects actually exist, exhausting all (actually) possible extensions does not produce first-order logic. Instead, this produces 'constant-domain first-order logic', the usual formal definitions relative to a single, unchanging domain of (actually existing) objects.[9] The simplest differences introduced by the use of a constant domain concern the existence of objects themselves. In first-order logic, $\exists x \exists y(x \neq y)$ is a contingent claim; there exist models which make this claim true and models which make it false. In constant-domain first-order logic, however, $\exists x \exists y(x \neq y)$ is a logical truth; there actually exists two objects, and so every model contains at least two objects. First-order logic is thus committed to strictly more than just appropriate extensions for each part of a first-order language.

As the discrepancy above suggests, the additional commitment built into first-order logic concerns objects themselves. First-order logic properly endorses both a normative claim about first-order languages and their extensions as well as a further normative claim about how objects *could be*, viz. any set whatever. There are a number of reasons this latter claim is attractive; the important point for current purposes, however, is only the modal nature of this assertion. First-order logic is not just swapping linguistic extensions. It imposes a real restriction on how the world could be.

In first-order logic proper, this restriction is only used to circumscribe possible models relative to a language. Characterizing possible domains with sets also unlocks, however, an exhaustive characterization of *atomic possibilities*. If sets exhaust all possible domains, then whatever the atomic structure of the world, there exists a corresponding set. Further, the extensions of any (first-order) properties must correspond to subsets of the domain set. Taking all possible collections of subsets of all possible sets thus delivers all atomic possibilities. By providing an exhaustive characterization of all the objects there could be, first-order logic likewise allows for an exhaustive characterization of all atomic possibilities.

4.5.2 Finite, Atomic Confirmation for First-Order Languages

The proposed derivation of equal possibility from atomic possibility also admits of a straightforward first-order formalization.

[9] Interestingly, Mancosu (2006, 2010) observes that Tarski's early logic papers use a constant domain semantics of precisely the sort suggested by this view.

Finite, Atomic Confirmation for First-Order Languages

Let L be a first-order language, $\Omega_{\mathcal{L}}$ a collection of canonical-domain models for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set of evidence.

(1') If $\Omega_{\mathcal{L}}$ is an exhaustive collection of atomic possibilities, then every possibility in the subcollection $\Omega_{\mathcal{L},\Sigma}$ is an equal possibility relative to $\Omega_{\mathcal{L}}$ and Σ.

(2') For a finite, non-empty set of equal possibilities, the degree of confirmation for a sentence φ is the number of equal possibilities which make φ true divided by the total number of equal possibilities.

Talk of possibilities is again understood in terms of models for some formal language, and degrees of confirmation result from taking a finite set of these models to be atomic.

Canonical-Domain Models

The first major complexity introduced by the move to first-order atomic confirmation is the notion of model to be used. A minimal example shows the insufficiency of standard first-order models for the proposed account:

> *Small World 1.*
> Suppose that the world contains either only a single object or exactly two objects. Formally, the first-order sentence $\forall x \forall y \forall z(x = y \vee y = z \vee x = z)$ is true. Moreover, the world is—as it turns out—quite bland; the only non-trivial property in the world is the single non-identity just noted. How many total possibilities exist here? How many of these ways make it the case that there exists exactly one object, i.e., that $\exists! x(x = x)$ is true?

The intuitive answers are two and one respectively. The world contains either one object or two, and only the former makes $\exists! x(x = x)$ true. In contrast with this first,

> *Small World 2.*
> Suppose again that the world contains either a single object or two, but in addition a single object in the world has been stamped with a blue mark. How many total possibilities exist? How many of these ways make it the case that there exists exactly one object, i.e., that $\exists! x(x = x)$ is true?

The intuitive answers are in this case three and one respectively. If there is only a single object, it must be the one stamped blue; if there are two objects, then either one or the other may be stamped. $\exists! x(x = x)$ is, of course, only true on the first of these ways.

The existence of a blue mark in *Small World 2* shifts the answers to our two questions by introducing additional structure for the world. Even if we do not know which entity bears the mark, some object has acquired an additional facet to its existence, and the space of possibilities must shift accordingly. The problem with first-order models is that they do not recognize shifts of this kind. There are only two (non-isomorphic) first-order models which satisfy the constraints stipulated in *Small World 2*:

$$\mathcal{M}_1 : \{a_0\} \quad b^{\mathcal{M}_1'} = a_0 \qquad\qquad \begin{array}{l} \mathcal{M}_2 : \{a_0, a_1\} \\ \qquad b^{\mathcal{M}_2'} = a_0 \\ \qquad \cong \\ \mathcal{M}_3 : \{a_0, a_1\} \\ \qquad b^{\mathcal{M}_2'} = a_1 \end{array}$$

More generally, first-order models are only regarded as distinct when no extension-preserving bijection exists between their respective domains. All individual objects are, on this approach, interchangeable. This suffices with the all or nothing nature of logical consequence but badly miscounts with the finer-grained notion of confirmation.

From a purely objective perspective, objects are distinct regardless of whether or not this fact is or even could be recognized. There is thus a difference between a blue dot on object a_0 and a blue dot on object a_1 in *Small World 2* even if we should lack the resources to actually distinguish a_0 from a_1. God, as it were, knows the difference even if we do not. First-order models ignore this and conflate distinct possibilities on the basis that they cannot be distinguished using only the supplied extensions. As a result, the first-order equipossibility account requires a revised notion of model which distinguishes objects even in the absence of extensional differences. Canonical-domain models accomplish precisely this by building an explicit labeling over objects into the models themselves. Relative to a fixed first-order language, first-order models supply all possibilities which can be distinguished by their extensions; canonical-domain models supply all possibilities simpliciter.

While only the notion of model deviates from the standard construction of first-order logic, a complete set of definitions for the desired logic is provided alongside the proofs for this section in the appendix. Underpinning the definition of a canonical-domain model, we will need a collection of canonical domains.

Definition 4.8 A *canonical-domain class* C is a class of sets satisfying

(i) for every cardinality, there exists a unique set of that cardinality within C

(ii) for any two cardinalities $\aleph_i$ and $\aleph_{i'}$ such that $\aleph_i \leq \aleph_{i'}$, the set of cardinality $\aleph_i$ is a subset of the set of cardinality $\aleph_{i'}$

A canonical-domain class is a collection of nested sets that contains exactly one set for every cardinality. The particular choice of a canonical-domain class C is irrelevant; we merely wish to secure a canonical set at each cardinality. For ease of use, we may suppose that elements of the smaller sets are subscripted by natural and then real numbers. The unique one element domain set is thus $\{a_0\}$, the unique two element domain set is $\{a_0, a_1\}$, and so forth.

Unlike first-order models, canonical-domain models all draw their domain from a particular canonical-domain class C.

Definition 4.9 For a first-order language L, a *canonical-domain L-model* $\mathcal{M}$ is a canonical domain dom($\mathcal{M}$) $\in C$ ("the domain set") together with an extension drawn from the domain set for each component of L's signature, i.e.,

- for any constant-symbol c_i, $\mathcal{M}$ supplies an element a_i from the domain set, $c_i^{\mathcal{M}}$.
- for any n-ary function-symbol f_i, $\mathcal{M}$ supplies an n-ary function over the domain set, $f_i^{\mathcal{M}}$.
- for any n-ary relation symbol R_i, $\mathcal{M}$ supplies an n-ary relation over the domain set, $R_i^{\mathcal{M}}$.

Examples of canonical-domain models include each of $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ above. Further, these last two now represent distinct models.

Having defined a collection of models over first-order languages, a model-theoretic consequence relation immediately follows.

Definition 4.10 For a first-order language L, $\Gamma \subseteq L$, and $\varphi \in L$, φ is a *canonical-domain consequence* of Γ, notated $\Gamma \models_C \varphi$, if and only if every canonical-domain model which makes Γ true also makes φ true.

This relation is equivalent to that induced by the standard definitions.

Proposition 4.10

Let L be a first-order language, $\Gamma \subseteq L$, *and* $\varphi \in L$. *Then,*

$$\Gamma \models \varphi \text{ if and only if } \Gamma \models_C \varphi.$$

The modified notion of model used here thus fails to introduce any change in the resulting consequence relation. More generally, for any canonical-domain model there exists an isomorphic first-order model and vice-versa; the difference between the two approaches is restricted only to equivalence between models. In this area, the canonical-domain approach recognizes strictly more distinctions than the standard construction as a direct result of tracking distinct objects.

The First-Order Atomic Confirmation Ranking

Given a first-order language L, a finite set of atomic canonical-domain models $\Omega_{\mathcal{L}}$ defined over L, and evidence $\Sigma \subseteq L$ consistent with $\Omega_{\mathcal{L}}$, the first-order atomic confirmation account supplies a unique degree of confirmation for every sentence in L.

Definition 4.11 For a first-order language L, finite set of canonical-domain models $\Omega_{\mathcal{L}}$ for L, and evidence $\Sigma \subseteq L$ consistent with $\Omega_{\mathcal{L}}$, the *first-order atomic confirmation ranking* is

$$\mu_{\Sigma}^{\mathcal{L}}(\varphi) = \frac{|[\![\varphi]\!]_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

Definition 4.12 For a first-order language L, finite set of canonical-domain models $\Omega_{\mathcal{L}}$ for L, and evidence $\Sigma \subseteq L$ consistent with $\Omega_{\mathcal{L}}$, the *comparative first-order atomic confirmation ranking* is

$$\varphi \precsim_{\Sigma}^{\mathcal{L}} \psi \Leftrightarrow \mu_{\Sigma}^{\mathcal{L}}(\varphi) \leq \mu_{\Sigma}^{\mathcal{L}}(\psi).$$

As with the propositional account, L together with a set of models $\Omega_{\mathcal{L}}$ for L straightforwardly gives rise to the algebras $\mathcal{F}_{\mathcal{L}} = \{[\![\varphi]\!]_{\Omega_{\mathcal{L}}} \ : \ \varphi \in L\}$ and $\mathcal{F}_{\mathcal{L}}^{+} = \mathcal{P}(\Omega_{\mathcal{L}})$. Further, both $\mu_{\Sigma}^{\mathcal{L}}$ and $\precsim_{\Sigma}^{\mathcal{L}}$ may be unambiguously extended to $\mathcal{F}_{\mathcal{L}}^{+}$ by setting

$$\mu_{\Sigma}^{\mathcal{L}}(E) = \frac{|E \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}$$

for any $E \in \mathcal{F}_{\mathcal{L}}^{+}$.

Just like finite atomic confirmation over propositional languages, finite atomic confirmation for first-order languages is not only probabilistic but Bayesian.

Proposition 4.11

Let L be a first-order language, $\Omega_{\mathcal{L}}$ a finite set of canonical-domain L-models defined over L, and $\Sigma \subseteq L$ a set of sentences consistent with $\Omega_{\mathcal{L}}$. Then, $\mu_{\Sigma}^{\mathcal{L}}$ satisfies Kolmogorov's probability axioms over $\mathcal{F}_{\mathcal{L}}^{+}$.

Corollary 4.1

Let L be a first-order language, $\Omega_{\mathcal{L}}$ a finite set of canonical-domain L-models defined over L, and $\Sigma \subseteq L$ a set of sentences consistent with $\Omega_{\mathcal{L}}$. Then, for any φ consistent with $\Omega_{\mathcal{L},\Sigma}$ and any ψ,

$$\mu_{\Sigma}^{\mathcal{L}}(\psi|\varphi) = \mu_{\Sigma\cup\{\varphi\}}^{\mathcal{L}}(\psi).$$

The extension to first-order languages thus shifts none of the high-level properties of the account.

4.5.3 The Reemergence and Resolution of D'Alembert's Riddle

As with the propositional account, the finite atomic confirmation ranking for first-order languages exhibits a general sensitivity to interpreted language extensions. The move to a first-order account alone thus blocks neither d'Alembert's riddle itself nor the underlying phenomenon. The move to a first-order account does, however, provide the resources for a principled response.

The Reemergence of D'Alembert's Riddle

The earlier discussion of interpreted propositional languages carries over almost verbatim to the case of first-order languages. As earlier, we restrict our attention in this subsection to model spaces $\Omega_{\mathcal{L}}$ that are composed entirely of L-models.

Definition 4.13 An *interpreted first-order language* $\mathcal{L}$ is a pair of a first-order language L and a non-empty collection $\Omega_{\mathcal{L}}$ of canonical-domain L-models.

The notion of an interpreted language extension is likewise unchanged from the propositional account.

Definition 4.14 A first-order interpreted language $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+}\rangle$ *extends* a first-order interpreted language $\mathcal{L} = \langle L, \Omega_{\mathcal{L}}\rangle$ if and only if $L \subseteq L^+$ and for any model $\mathcal{M}^+$ in $\Omega_{\mathcal{L}^+}$, $\mathcal{M}^+|_L \in \Omega_{\mathcal{L}}$.

Finally, the behavior of the atomic confirmation ranking with respect to different interpreted languages can again be usefully characterized by restricting our attention to extensions.

Formalizing the small world example from the previous section establishes that the first-order confirmation account is sensitive to extensions of the underlying interpreted language. In this case, an empty signature is expanded by a single constant b.

Small World 1		Small World 2
		$\mathcal{M}'_1 : \{a_0\}$, $b^{\mathcal{M}'_1} = a_0$
$\mathcal{M}_1 : \{a_0\}$		$\mathcal{M}'_2 : \{a_0, a_1\}$, $b^{\mathcal{M}'_2} = a_0$
$\mathcal{M}_2 : \{a_0, a_1\}$	$\Rightarrow$	$\mathcal{M}'_3 : \{a_0, a_1\}$, $b^{\mathcal{M}'_3} = a_1$

By adding a new constant b to the underlying first-order language, some object has been (metaphorically) emblazoned with a special mark. Not only do the possibilities under consideration shift in response but so too does the degree of confirmation attributed to the sentence $\exists! x(x = x)$.

Just as with the propositional case, extensions of interpreted first-order languages divide into three general categories. Unrestricted extensions of an interpreted language impose no restrictions on the extensions of the introduced symbols.

Definition 4.15 A first-order interpreted language $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+}\rangle$ is an *unrestricted extension* of a first-order interpreted language $\mathcal{L} = \langle L, \Omega_{\mathcal{L}}\rangle$ if and only if $\mathcal{L}^+$ is an extension of $\mathcal{L}$ and $\Omega_{\mathcal{L}^+}$ contains every canonical-domain model $\mathcal{M}^+$ satisfying $\mathcal{M}^+|_L \in \Omega_{\mathcal{L}}$.

Unrestricted extensions thus occur when moving from a starting interpreted language $\mathcal{L}$ to an interpreted language $\mathcal{L}^+$ with no accompanying limits on the resulting models. The example below illustrates an unrestricted extension by a unary relation-symbol P:

$$
\begin{array}{ccl}
 & & \mathcal{M}'_1 : \{a_0\} \quad P^{\mathcal{M}'_1} = \{\} \\
 & & \mathcal{M}'_2 : \{a_0\} \quad P^{\mathcal{M}'_2} = \{a_0\} \\
\mathcal{M}_1 : \{a_0\} & & \mathcal{M}'_3 : \{a_0, a_1\} \quad P^{\mathcal{M}'_3} = \{\} \\
 & \begin{array}{c}\text{Unrestricted}\\ \text{Extension}\\ \Rightarrow\end{array} & \\
\mathcal{M}_2 : \{a_0, a_1\} & & \mathcal{M}'_4 : \{a_0, a_1\} \quad P^{\mathcal{M}'_4} = \{a_0\} \\
 & & \mathcal{M}'_5 : \{a_0, a_1\} \quad P^{\mathcal{M}'_5} = \{a_1\} \\
 & & \mathcal{M}'_6 : \{a_0, a_1\} \quad P^{\mathcal{M}'_6} = \{a_0, a_1\}
\end{array}
$$

Note that the unrestricted language extension allows for models with every possible extension for P relative to the original model space. This kind of language extension corresponds to the introduction of a new symbol with no antecedent connection to other parts of the language.

Unlike this first kind, a completely restricted extension of an interpreted language dictates a unique extension for each added symbol.

Definition 4.16 An interpreted first-order language $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+}\rangle$ is a *completely restricted extension* of a first-order interpreted language $\mathcal{L} = \langle L, \Omega_{\mathcal{L}}\rangle$ if and only if $\mathcal{L}^+$ is an extension of $\mathcal{L}$ and for every canonical-domain model $\mathcal{M}$ in $\Omega_{\mathcal{L}}$, there exists a unique canonical-domain model $\mathcal{M}^+$ in $\Omega_{\mathcal{L}^+}$ such that $\mathcal{M}^+|_L = \mathcal{M}$.

Thus, while we again transition from $\mathcal{L}$ to an extended language $\mathcal{L}^+$, the models of the extended language are in complete agreement as to the extension of the new symbol(s):

$$\begin{array}{lcl} \mathcal{M}_1 : \{a_0\} & & \begin{array}{l}\mathcal{M}_1'' : \{a_0\} \\ \quad P^{\mathcal{M}_1''} = \{a_0\}\end{array} \\ & \begin{array}{c}\text{Completely}\\ \text{Restricted}\\ \text{Extension}\\ \Rightarrow\end{array} & \\ \mathcal{M}_2 : \{a_0, a_1\} & & \begin{array}{l}\mathcal{M}_2'' : \{a_0, a_1\} \\ \quad P^{\mathcal{M}_2''} = \{a_0\}\end{array} \end{array}$$

Informally, the meaning of the new symbols entirely determines what their extensions are. We may, for example, imagine introducing a new symbol that picks out all and only *that object* or which by definition picks out nothing at all.

The third and final kind of extension are those intermediate between unrestricted and completely restricted.

Definition 4.17 An interpreted first-order language $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+}\rangle$ is a *partially restricted extension* of an interpreted first-order language $\mathcal{L} = \langle L, \Omega_{\mathcal{L}}\rangle$ if and only if $\mathcal{L}^+$ is an extension of $\mathcal{L}$ and $\mathcal{L}^+$ is neither an unrestricted extension nor a completely restricted extension.

Extensions of this sort neither dictate a unique extension for the new symbols nor allow all possible extensions. The meaning of the new symbols is partially restricted. Consider, for example, adding a new predicate which must be non-empty.

$$
\begin{array}{lcl}
 & & \mathcal{M}_1'' : \{a_0\} \\
 & & \qquad P^{\mathcal{M}_1''} = \{a_0\} \\
 & & \\
 & & \mathcal{M}_2'' : \{a_0, a_1\} \\
\mathcal{M}_1 : \{a_0\} & & \qquad P^{\mathcal{M}_2''} = \{a_0\} \\
 & \text{Partially} & \\
 & \text{Restricted} & \\
 & \text{Extension} & \mathcal{M}_3'' : \{a_0, a_1\} \\
\mathcal{M}_2 : \{a_0, a_1\} & \Rightarrow & \qquad P^{\mathcal{M}_3''} = \{a_1\} \\
 & & \\
 & & \mathcal{M}_4'' : \{a_0, a_1\} \\
 & & \qquad P^{\mathcal{M}_4''} = \{a_0, a_1\}
\end{array}
$$

The extension of P is neither fixed nor allowed to vary freely.

As in the propositional case, completely restricted extensions of interpreted first-order languages fix the atomic confirmation ranking.

Proposition 4.12

Let $\mathcal{L}$ be an interpreted first-order language, $\mathcal{L}^+$ a completely restricted extension of $\mathcal{L}$, and $\Sigma \subseteq L$. If $\Omega_{\mathcal{L},\Sigma}$ is non-empty and finite, then for any $A \in \mathcal{F}^+_{\mathcal{L}}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{\mathcal{M}^+ \in \Omega_{\mathcal{L}^+} \ : \ \mathcal{M}^+|_L \in A\}$.

If $A = [\![\varphi]\!]_{\Omega_{\mathcal{L}}}$ for some $\varphi \in L$ above, we thus have $\mu^{\mathcal{L}}_{\Sigma}(\varphi) = \mu^{\mathcal{L}^+}_{\Sigma}(\varphi)$. In the first-order case, unrestricted extensions no longer need preserve the atomic confirmation ranking. Both the example of an unrestricted extension provided earlier and the small world example illustrate this. Only when restricted to a fixed domain is preservation guaranteed.

Proposition 4.13

Let $\mathcal{L}$ be a first-order interpreted language, $\mathcal{L}^+$ an unrestricted extension of $\mathcal{L}$ by a finite number of symbols, and $\Sigma \subseteq L$. If every canonical-domain model $\mathcal{M} \in \Omega_{\mathcal{L}}$ has the same domain and $\Omega_{\mathcal{L},\Sigma}$ is both nonempty and finite, then for any set $A \in \mathcal{F}^+_{\mathcal{L}}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{\mathcal{L}^+\text{-models } \mathcal{M}^+ \in \Omega_{\mathcal{L}^+} \ : \ \mathcal{M}^+|_L \in A\}$.

Finally, partially restricted extensions only occasionally preserve the atomic confirmation ranking. Examples here have the air of happenstance; multiple additions which alone would produce a shift but when combined balance one another out. The atomic confirmation ranking for first-order languages thus exhibits a widespread sensitivity to the particular interpreted language used. D'Alembert's riddle extends here too.

Resolving D'Alembert's Riddle

The first-order atomic confirmation account endorses a widespread dependence between confirmation and interpreted first-order languages. In the presentation of d'Alembert's riddle, this sort of dependence was cast as a severe loss. This subsection revisits the topic from a first-order perspective and argues that (1) a unique collection of atomic models can, in fact, be identified for any well-defined set of evidence and (2) confirmation ought to be sensitive to the exact set of atomic models under consideration. Both components of this defense derive from the interrelation of atomic canonical-domain models and properties.

By a 'property', I understand a feature of the world itself (as opposed to either an extension or a property-name). This is some bit of structure that is stamped into reality rather than a product of oneself or a component of a useful model. Properties in this sense underwrite much of our talk of extensions but are not themselves only sets. I take it that upon reflection it is obvious that properties permeate our world view. We really do think that there are such things as being green, being gold, and being a father. These are something independent of us and beyond mere collections of disparate objects. It is of course possible to deny that this or that is a property in my intended sense. Regardless, the use of properties itself—like the use of truth or causation—is a tenacious practice, and I adopt it here without further defense.

Properties are particularly important in the current framework because adopting a particular collection of canonical-domain models as atomic entails a corresponding position on the number and kind of properties in the world. This interrelation between atomic models and properties is most prominent in unrestricted extensions of an interpreted first-order language. Consider, for example, the unrestricted extension of an interpreted first-order language $\mathcal{L}$ by a single unary relation-symbol P. The strictures on first-order languages require that P, once introduced, has a definite extension. In the case of a first-order language, this is the extent of the change. Interpreted first-order languages, however, are equipped with a collection of atomic models. An unrestricted interpreted language extension entails not only a new symbol but also that P's extension varies freely across the collection of atomic models. Meaning alone does not determine when objects are in P and when they are not. P's extension must therefore be determined by the world itself. The relation-symbol P corresponds to a property stamped into the possibilities under consideration. This unrestricted extension of an interpreted language has more likeness to suddenly being informed that many of the objects which surround us are emblazoned with a glowing 'P' than to the mere addition of a new symbol to our vocabulary.

These comments extend well beyond the example above. In every unrestricted extension of an interpreted first-order language, the extensions of the introduced symbols cannot by their very nature be attributed solely to the meanings of the symbols in question and so too must correspond to properties. Partially restricted extensions of an interpreted first-order language provide a mixed case. The extensions of symbols introduced in this manner are partially determined by their connection to other parts of the language. Properties must again be posited, however, in order to make up the shortfall between a partially determined extension and a determinate one. Indeed, only completely restricted extensions of an interpreted first-order language need not invoke properties. In this case, the extensions of new symbols are entirely determined by their relationship to other parts of the language. In the current framework, changes in interpreted first-order languages thus generally accompany revisions in the number and kind of properties accepted.

In light of this, the atomic confirmation ranking's sensitivity to the particular interpreted first-order language adopted takes on a new significance. Since the only kind of interpreted language extension which doesn't introduce new properties likewise fixes the atomic confirmation ranking, confirmation disagreements between an interpreted language and some extension thereof reduce to a disagreement over properties. Noting that any two first-order languages L and L' can be viewed as extensions of a common language L^- and that this in turn is an extension of the language with no signature $L_\emptyset$, confirmation disagreements between any two interpreted first-order languages must derive either from an antecedent disagreement about how many objects there are (disagreement at the $L_\emptyset$ stage), a disagreement over the meaning of their common symbols (disagreement at L^- stage), or a disagreement over properties (disagreement at the L, L' stage). In all three cases, the confirmation relation intuitively should vary between the two perspectives.

Dependence on the particular collection of atomic models adopted in the first-order atomic confirmation account is thus necessary. Differences in these collections correspond to divergent positions on the meaning of symbols, on the structure of the world itself, or a combination of the two. The choice of atomic models is thus neither arbitrary nor inconsequential. Different selections represent materially different views, and a precise degree of confirmation requires as it were a precise question.

Returning to d'Alembert's riddle, the two different solutions here are the result of two contrasting views about the properties surrounding coins. The standard solution holds that the world demands the equivalence of not landing heads and landing tails. In contrast, d'Alembert's proposal maintains that landing heads on the first toss necessitates no second toss. These are two different pictures of the world and adopting one or the other naturally shifts the degree of confirmation for a coin landing heads within two tosses. Intuitions in this case are further complicated by the fact that both these solutions are absurd. Not only are the models here clearly not atomic, but our explicit evidence demands neither the equivalence of not landing heads and landing tails nor that landing heads on the first toss necessitates no second toss. The interpreted language sensitivity displayed by d'Alembert's riddle is important but both proposed solutions to the riddle are inadequate.

The beginning of this section noted that first-order logic supposes that sets exhaust all possible domains and from this a characterization of all atomic possibilities could be had. This characterization was then developed in the following subsection in the form of canonical-domain models. Whatever the atomic structure of the world really is, there exists a corresponding canonical-domain model in some language. From this perspective, adopting a particular collection of atomic models operates as a particular sort of evidence, viz. that there are these properties and they are related like so. The standard solution and d'Alembert's proposal, for example, are naturally construed as making two different assumptions about how the world is structured. Generalizing away from a particular collection of atomic models is then only a matter of reversing course and including more canonical-domain models for more first-order languages.

At the furthest extreme, we have the situation raised by d'Alembert's riddle. Given only a set of evidence Σ in some first-order language L, can we identify an objective collection of atomic models? The answer is that we can. The collection we desire is the collection of all canonical-domain models (for either L or an extension of L) that make Σ true. If the collection of all canonical-domain models gives all atomic possibilities, then the collection of all canonical-domain models consistent with Σ provides all atomic possibilities consistent with the evidence Σ. As a result, the proposed account does not need to suppose that a collection of atomic models is provided.

FINITE CONFIRMATION FOR FIRST-ORDER LANGUAGES

Let L be a first-order language and $\Sigma \subseteq L$ a consistent[10] set of evidence.

(0) The collection of all canonical-domain models is an exhaustive collection of atomic possibilities.

(1') If $\Omega_{\mathcal{L}}$ is an exhaustive collection of atomic possibilities, then every possibility in the subcollection $\Omega_{\mathcal{L},\Sigma}$ is an equal possibility relative to $\Omega_{\mathcal{L}}$ and Σ.

(2') For a finite, non-empty set of equal possibilities, the degree of confirmation for a sentence φ is the number of equal possibilities which make φ true divided by the total number of equal possibilities.

In the case where $\Omega_{\mathcal{L}}$ is the collection of all canonical-domain models, we will say that the subcollection $\Omega_{\mathcal{L},\Sigma}$ is the collection of equal possibilities relative only to Σ. Unless Σ itself is quite restrictive, the result of failing to provide an interpreted first-order language is a proliferation of atomic possibilities.

In the case of d'Alembert's riddle, the collection of all atomic possibilities consistent with the description of d'Alembert's game contains models with both arbitrarily large domains and arbitrarily many properties. Thus while d'Alembert's challenge has been met,

[10]If $\Omega_{\mathcal{L}}$ is supplied, this is consistency with respect to $\Omega_{\mathcal{L}}$. If $\Omega_{\mathcal{L}}$ is not supplied, this is first-order consistency or, equivalently, consistency with respect to the collection of all canonical-domain models.

the correct degree of confirmation for ‘the coin lands heads within two tosses’ requires an extension of (2’) beyond finite sets of equal possibilities. This challenge is taken up in the next chapter.

Chapter 5

A Logical Theory of Confirmation

This chapter extends the account of confirmation proposed in Chapter 4 to infinite collections. The primary barrier here is Bertrand's paradox, the last of our three objections to confirmation. The first section begins with a review of well-known instances of Bertrand's paradox as well as proposed solutions. Comparing these with the standard analysis of the paradox makes it clear, however, that commentators have systematically underestimated the phenomenon in question. The final contribution of the first section is thus a revised analysis of Bertrand's paradox that significantly expands its scope.

Even under this revised analysis, however, Bertrand's paradox fails to block accounts of confirmation generally. The second section of this chapter develops an extension of the finite confirmation account from chapter four that is both immune to the paradox and satisfies a large number of intuitive properties. This solution to Bertrand's paradox is, moreover, maximal; any alternative account of confirmation either falls victim to the paradox or is strictly weaker. Our discussion of confirmation ends with a return to absolute rankings and an extension of the proposed account to support talk about degrees of confirmation.

5.1 Bertrand's Paradox

Cataloged and popularized by Bertrand (1889), instances of Bertrand's paradox first appear in mathematics periodicals two decades earlier (Myrvold 2019). The most well known instances of the paradox are Bertrand's (1889) *Circles and Chords*, von Kries' (1886) *Specific Volume and Density*, von Mises' (1957) *Water and Wine*, and most recently van Fraassen's (1989) *The Mystery Cube Factory*. After presenting each variant, we also review both standard solutions to and the standard analysis of the paradox. Ironically, it is only Bertrand's (1889) *Circles and Chords* which fails to conform to the standard analysis. In light of this failure, a revised analysis of Bertrand's paradox is presented and defended.

5.1.1 Standard Instances of Bertrand's Paradox

Circles and Chords

Although Bertrand presented his paradox in a number of different settings, his most well-known formulation is geometric. Bertrand begins with an apparently well-defined question:

> Given a circle of radius r, how likely is it that an arbitrary[1] chord on this circle will be longer[2] than the side of an inscribed equilateral triangle?

A *chord* on a circle is a line segment each of whose endpoints lie on the circle; an example is pictured in Figure 1 below. The side-length of the unique (up to rotation) equilateral triangle which can be inscribed in a circle with radius r is $\sqrt{3}r$, pictured in Figure 2 below.

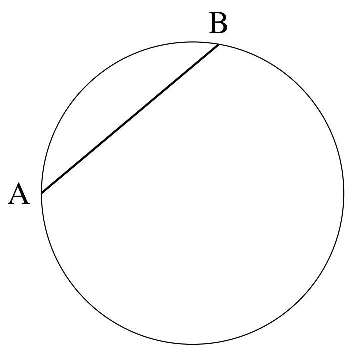


Figure 1

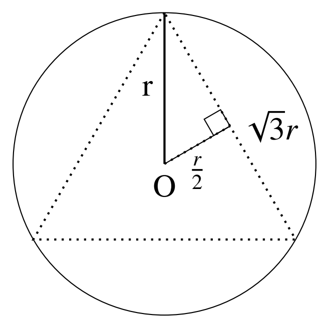


Figure 2

Bertrand proposes three methods by which the desired likelihood could be found. Each proposal leverages a small amount of geometry in order to characterize both all available chords and chords with length greater than $\sqrt{3}r$ in a more intuitively tractable format.[3]

[1] In Bertrand's original writings, he writes of taking a chord "au hasard" which is typically translated as "at random". I have substituted 'arbitrary' instead to highlight the availability of several distinct interpretations for both this challenge and the paradox that follows. If, for example, 'arbitrary' is interpreted as code for a particular distribution of physical propensities, then the resulting paradox pertains to physical propensities of this form. If 'arbitrary' is interpreted instead as shorthand for a particular distribution in the mathematical theory of probability (e.g., the uniform distribution), then the paradox pertains to the mathematical theory of probability. If 'arbitrary' is interpreted only as ruling out additional evidence about the chord selected, then the paradox pertains to absolute confirmation. Contemporary philosophy recognizes both a number of distinct notions of likelihood and a number of formal representations for these notions. Bertrand's paradox proper belongs to no single notion or formalism.

[2] Bertrand (1889) actually asks after chords shorter than the side of an inscribed equilateral triangle. Contemporary discussion of the paradox, however, almost exclusively concerns chords that are longer, and so I too adopt this formulation of the paradox.

[3] Shackel (2007) and Rowbottom (2013) rightly observe that there is cause for worry with all three of Bertrand's proposals. Each of the proposed methods treats a subset of the available chords and then lifts this treatment to the total space. No guarantee of the adequacy of such a method is ever explicitly provided, and in some cases the operation is highly unintuitive, e.g., in the third method, the centerpoint of the circle corresponds with uncountably many chords while other points correspond to only one (Rowbottom 2013).

Circles and Chords - Distance from Origin to Chord Midpoint

A chord $\overline{AB}$ is longer than $\sqrt{3}r$ if and only if the distance from the origin O to the midpoint of $\overline{AB}$, labeled $M_{\overline{AB}}$ in Figure 3, is between 0 and $\frac{r}{2}$ (compare with the inscribed equilateral triangle in Figure 2). Since the length between O and $M_{\overline{AB}}$ can be anywhere in the interval $[0, r]$, exactly half of all the possible lengths result in a chord longer than $\sqrt{3}r$. Since $\overline{AB}$ was arbitrary, a chord longer than $\sqrt{3}r$ will then occur with likelihood $\frac{1}{2}$!

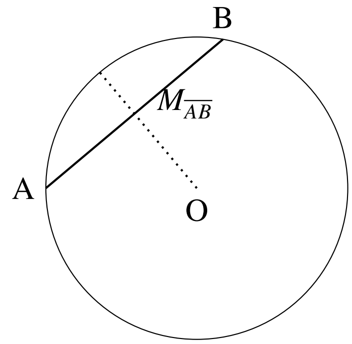


Figure 3

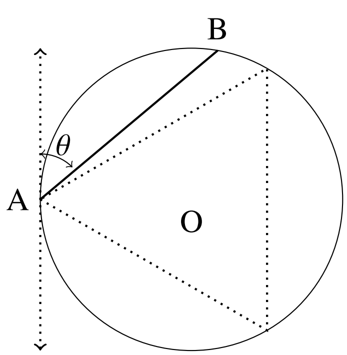


Figure 4

Circles and Chords - Angle from Tangent

For any point A, we can inscribe an equilateral triangle with A itself as one of the triangle's points. The endpoint B of the chord is entirely determined by the angle θ between A and the line tangent to the circle at A as in Figure 4. Since the sides of the inscribed equilateral triangle occur at 60° and then again at 120° from the tangent, the chord $\overline{AB}$ is longer than $\sqrt{3}r$ if and only if θ is strictly between 60° and 120°. Since θ can be drawn from anywhere in the interval $[0°, 180°)$, exactly one third of all possible angles result in a chord longer than $\sqrt{3}r$. A chord longer than $\sqrt{3}r$ will then occur with likelihood $\frac{1}{3}$!

Circles and Chords - Area of Midpoint Regions

Just as above, a chord $\overline{AB}$ is longer than $\sqrt{3}r$ if and only if the distance from the origin O to the midpoint of $\overline{AB}$ is less than $\frac{r}{2}$ (compare with Figure 2). Since any point within the circle C is the midpoint of some chord and only those points within the smaller circle C' of radius $\frac{r}{2}$ are midpoints of chords longer than $\sqrt{3}r$, we need only take the area of C' divided by the area of C:

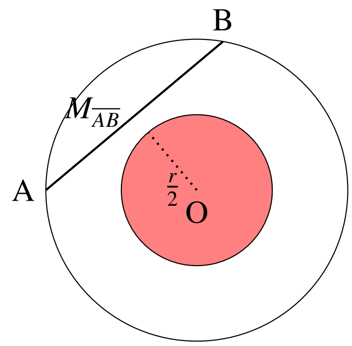


Figure 5

$$\begin{aligned}\frac{\text{Area}(C')}{\text{Area}(C)} &= \frac{\pi(\frac{r}{2})^2}{\pi r^2} \\ &= \frac{\frac{1}{4}r^2}{r^2} \\ &= \frac{1}{4}.\end{aligned}$$

A chord longer than $\sqrt{3}r$ will then occur with likelihood $\frac{1}{4}$!

Each of Bertrand's proposed methods thus generates a distinct value. Since the requested likelihood cannot be simultaneously $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$, at most one of these methods is legitimate.

Specific Volume and Density

The German logician von Kries offers his own variant of the paradox. Unlike Bertrand, von Kries (1886) formulates the paradox using two intrinsic properties of matter: specific volume and density. The density of an object is the ratio of the object's mass to its volume; the specific volume of an object meanwhile is the ratio of its volume to its mass. Consider then the following two cases:

> *Specific Volume and Density - Specific Volume*
> A substance is presented which is known only to have specific volume between 1 and 3. How likely is it that the substance's specific volume is between 1 and 2? How likely is it that the substance's specific volume is between 2 and 3?

The intuitive answer is that the requested intervals each represent half of the total collection of possibilities—the entire $[1, 3]$ interval—and so the desired likelihoods are each $\frac{1}{2}$. Consider now a similar situation with density:

> *Specific Volume and Density - Density*
> A substance is presented which is known only to have density between $\frac{1}{3}$ and 1. How likely is it that the substance's density is between $\frac{1}{3}$ and $\frac{2}{3}$? How likely is it that the substance's density is between $\frac{2}{3}$ and 1?

Again, the intuitive answer is that the requested intervals each represent half of all possible density values—in this case, the interval $[\frac{1}{3}, 1]$—and so the desired likelihoods are both again $\frac{1}{2}$. Finally, note that the density of a substance is simply the reciprocal of its specific volume. Thus, if the substance has specific volume v, it necessarily has density $\frac{1}{v}$. The two situations described above are thus equivalent to one another, and the intuitive answers supplied should cohere. Supposing that there is a non-zero likelihood for any possible interval, these answers are instead inconsistent. If a specific volume between 1 and 2 has likelihood $\frac{1}{2}$, then so too must a density between $\frac{1}{2}$ and 1. This is only consistent with densities between $\frac{2}{3}$ and 1 also having likelihood $\frac{1}{2}$ however if densities between $\frac{1}{2}$ and $\frac{2}{3}$ have likelihood zero—which is intuitively not the case. Despite the correspondence between specific volume and density, the intuitive solution thus shifts based on which property of the substance is emphasized.

The contradictory commitments here can be made more obvious with a minor modification to the original puzzle. Consider again the specific volume formulation of the paradox, only instead we desire the likelihood that this lies in the interval $[1, \frac{3}{2}]$. Since the specific volume must be drawn from $[1, 3]$, the requested likelihood is intuitively $\frac{1}{4}$. A density between $[\frac{2}{3}, 1]$ meanwhile intuitively makes up $\frac{1}{2}$ the total space of available densities, $[\frac{1}{3}, 1]$. Since specific volumes in $[1, \frac{3}{2}]$ correspond to densities in $[\frac{2}{3}, 1]$, the same possibility has been assigned both likelihood $\frac{1}{4}$ and $\frac{1}{2}$—a contradiction.

Water and Wine

Richard von Mises (1957) offers the following variation:

> *Water and Wine - Water/Wine*
> A glass is presented which contains a mixture of water and wine. It is known only that the glass contains at least as much water as wine and at most twice as much water as wine, i.e., the proportion $\frac{\text{water}}{\text{wine}}$ is in $[1, 2]$. How likely is it that the ratio of water to wine is between 1 and $\frac{3}{2}$? How likely is it that the ratio of water to wine is between $\frac{3}{2}$ and 2?

The intuitive answer is that the requested intervals each represent half of the total collection of possibilities—the entire $[1, 2]$ interval—and so the desired likelihoods are each $\frac{1}{2}$. Consider now the exact same situation only rephrased in terms of the ratio of wine to water:

> *Wine and Water - Wine/Water*
> A glass is presented which contains a mixture of water and wine. It is known only that the glass contains at least half as much wine as water and no more than an even mix, i.e., the proportion $\frac{\text{wine}}{\text{water}}$ is in $[\frac{1}{2}, 1]$. How likely is it that the ratio of wine to water is between $\frac{1}{2}$ and $\frac{3}{4}$? How likely is it that the ratio of wine to water is between $\frac{3}{4}$ and 1?

Again, the intuitive answer is that the requested intervals each represent half of all possible values—in this case, the interval $[\frac{1}{2}, 1]$—and so the desired likelihoods are both again $\frac{1}{2}$. Finally, observe that water to wine ratios are just the reciprocals of wine to water ratios and vice-versa. If we suppose, as von Mises (1957) appears to,[4] that there is a non-zero likelihood for any possible interval, then the intuitive answers in the two cases above are inconsistent with one another. A water to wine ratio between 1 and $\frac{3}{2}$ is a wine to water ratio between $\frac{2}{3}$ and 1; this is only consistent with our second set of responses if wine to water ratios between $\frac{2}{3}$ and $\frac{3}{4}$ are assigned likelihood zero.

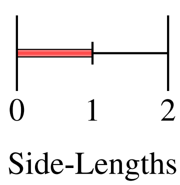
A more direct contradiction can be established by asking after water and wine mixtures in the interval $[1, \frac{4}{3}]$. Since the mixture must lie somewhere within $[1, 2]$, the likelihood of a mixture in $[1, \frac{4}{3}]$ is intuitively $\frac{1}{3}$. Considering instead ratios of wine to water, all values in $[\frac{1}{2}, 1]$ are possible. The likelihood that the wine to water ratio is in $[\frac{3}{4}, 1]$ is intuitively $\frac{1}{2}$. Since water to wine ratios and wine to water ratios are reciprocals, the two intuitive solutions of $\frac{1}{3}$ and $\frac{1}{2}$ directly contradict one another.

The Mystery Cube Factory

The best-known instance of Bertrand's paradox is *The Mystery Cube Factory* first elaborated by van Fraassen (1989):

> *The Mystery Cube Factory - Length.*
> A factory produces perfect cubes with side-length ≤ 2 cm. Given this, how likely is it that a cube produced by this factory has side-length ≤ 1 cm?

The expected answer is that the requested value is $\frac{1}{2}$. If cubes may have any side-length less than 2 cm, then those side-lengths less than 1 cm intuitively make up half the total interval.

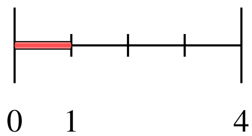


Of course, the description in terms of side-length above isn't necessary; the mystery cubes at issue also have areas, volumes, and so forth. Consider a similar question then with area instead of side-length:

> *The Mystery Cube Factory - Area.*
> A factory produces perfect cubes with faces whose areas are ≤ 4 cm^2. Given this, how likely is it that a cube produced by this factory has a face with area ≤ 1 cm^2?

[4] Richard von Mises (1957) claims that the intuitive attributions themselves are inconsistent; this is not the case. In von Mises' defense, proponents of (POI) are committed to a number of premises sufficient for a contradiction at this point.

The intuitive answer is now that the requested likelihood is $\frac{1}{4}$. If cubes may have any area less than 4 cm^2, then those with area less than 1 cm^2 intuitively make up a quarter of all possibilities.



The paradox is sealed by noticing that a cube with side-length less than 2 cm is a cube whose faces have area less than 4 cm^2. Similarly, a cube with side-length less than 1 cm is a cube whose faces have area less than 1 cm^2. The intuitive answers of $\frac{1}{2}$ and $\frac{1}{4}$ are thus inconsistent.

5.1.2 Standard Solutions to Bertrand's Paradox

Solutions to Bertrand's paradox have been largely skeptical in nature. Bertrand (1889) endorsed this reaction, writing:

> Among these three answers, which one is proper? None of the three is incorrect, none is correct, the question is ill-posed. (5)

While Bertrand ultimately favored a move to finitism, most commentators have embraced a less radical proposal. In so far as there exists a standard solution to the paradox, this solution holds that selecting a chord "au hasard" is simply ambiguous. For a definite solution to be identified, a more precise description of the desired procedure is required. Marinoff (1994) provides a well-known, contemporary instance of this line though similar sentiments were expressed by both Borel and Poincaré[5] decades earlier (Keynes 1921, 52-53).

This standard response successfully resolves the paradox from the perspective of contemporary probability theory but affords little protection to theories of confirmation. Even if the standard response is correct about the ambiguity of 'au hasard', theories of confirmation claim to transform evidence—even ambiguous evidence—into a rigorous, objective measure of evidential support. The purported ambiguity in Bertrand's question ought then be no special barrier, and a single, correct degree of confirmation is still owed. By their very nature, theories of confirmation cannot make use of the standard response to Bertrand's paradox.

[5]Contra van Fraassen (1989), neither Jaynes (1973) nor Poincaré (1923) fall easily into the group of those "who believed that the Principle of Indifference could be refined and sophisticated, and thus saved from paradox" (306). Both authors explicitly recognized not only the shortcomings of their invariance criteria but also the tension between their response and the principle of indifference.

The Invariance Program

The best-known strategy for identifying a unique solution to an instance of Bertrand's paradox rests on invariance. This strategy originates with Poincaré's (1923) observation that there exists a unique probability density function on the xy plane,

$$p(x, y) = (x^2 + y^2)^{-\frac{3}{2}},$$

which is invariant under both rotations and translations (Kendall and Morán 1963, 14-16). Poincaré himself appears to have read relatively little into this result; Bertrand did not, after all, stipulate invariance under rotation and translation of the xy plane as a part of *Circles and Chords*. Two decades later, Jeffreys (1967, 99-107) made a similar observation in the context of *Specific Volume and Density*. For a positive, real-valued parameter x, setting

$$p(x) \propto \frac{1}{x}$$

gives a prior density function invariant (up to the choice of a constant) under power transformations. Unlike Poincaré, however, Jeffreys regarded this family of improper[6] priors as objectively correct.

Jeffreys' (1967) discussion here has proven influential. The basis of *Specific Volume and Density* is the observation that if one has no information about specific volumes, then one likewise has no information about densities. The true "uninformative" prior ought then to be preserved by the transformation $T : (0, \infty) \to (0, \infty)$ defined by $T(x) = x^{-1}$. Further, there is nothing special here about the exponent -1. Any power transformation $T(x) = x^{\beta}$ for $\beta \in \mathbb{R} - \{0\}$ provides a well-behaved, bijective map from $(0, \infty)$ to $(0, \infty)$. Accordingly, the "uninformative" prior over $(0, \infty)$ must be preserved under power transformations generally. Since $p(x) \propto \frac{1}{x}$ is the only prior distribution over $(0, \infty)$ invariant under all power transformations, this must then be the correct "uninformative" distribution.

Bertrand's paradox is thus a product of the mistaken assumption—sometimes dubbed "Bayes' Postulate"—that the uniform distribution is the "uninformative" distribution. Jeffreys contends instead that the correct representation of complete ignorance varies depending on the nature of the parameter under consideration. The uniform distribution is, for instance, the correct representation of complete uncertainty over a finite space while $p(x) = \frac{1}{x}$ is the correct representation over $(0, \infty)$. By carefully considering invariance properties, we may identify the "uninformative" distribution for still further cases. Jeffreys (1967) thus introduced the search for invariant or uninformative priors into the foundations of Bayesian statistics, arguing that Bertrand's paradox could be resolved simply by identifying the appropriate "uninformative" distribution for each kind of parameter.

[6] An *improper prior* is a measure μ which is not normalized, i.e., does not set $\mu(\Omega) = 1$. In many cases, such a function will still produce a posterior probability function despite the lack of normalization in the prior. In these circumstances, the resulting posterior probability function is fixed for any proportional prior function. This is the apparent basis for Jeffreys' lack of concern over endorsing a family of proportional priors instead of a single unique prior.

Jeffreys' argument for $p(x) \propto \frac{1}{x}$ is insufficient as stated. Power transformations are not the only well-behaved, bijective maps from $(0, \infty)$ to $(0, \infty)$, and it is easy to construct examples for which Jeffreys "uninformative" prior over $(0, \infty)$ is not invariant.[7] Consider, for instance,

$$T(x) = \begin{cases} x & x \in (0, 1) \\ 2x - 1 & \text{otherwise} \end{cases}$$

which shifts the density function over $[1, \infty)$:

$$\begin{aligned} \frac{1}{T(x)}\, dT(x) &= \frac{2}{2x-1}\, dx \\ &\not\propto \frac{1}{x}\, dx. \end{aligned}$$

The disagreement here is not restricted to only the form of the probability density function. The weight ascribed to the interval $[1, 2]$, for example, shifts under the transformation T:

$$\begin{aligned} \int_1^2 \frac{1}{x}\, dx &= \ln(2) \\ &\approx .693 \end{aligned}$$

while

$$\begin{aligned} \int_1^2 \frac{1}{T(x)}\, dT(x) &= \int_1^2 \frac{2}{2x-1}\, dx \\ &= \ln(3) \\ &\approx 1.099. \end{aligned}$$

If Jeffreys' solution to Bertrand's paradox is to succeed, a principled reason for restricting the invariance requirement to only power transformations is required.

For his part, Jeffreys does appear to attempt such a justification:

> The point may be put in another way. If a parameter v is a dimensional magnitude and not a number, and we want to assess $P(dv|H)$, where H contains no information about v except that it is positive, this can only be of the form $Av^n\ dv$ where A and n are constants. For the ratio of two probabilities must be a number, which would not be satisfied if we took the first factor, say, as $\sin v$—the sine of a length means nothing. Nor could it be $e^{-\frac{v}{a}}$, where a is some constant of the same dimensions as v. For then it would assign a definite value to the ratio of the probabilities that v is less or greater than a. If, then, a is known, it contradicts the condition that we know nothing about v except its existence and that it lies between 0 and $+\infty$...The coefficient of dv must be

[7] As Robert et al. (2009) observes, Jeffreys appears to recognize this in a subsequent paragraph but maintains that $p(x) \propto \frac{1}{x}$ is nevertheless better than the Bayes-Laplace rule. This sudden pragmatic turn is both at odds with his explicitly objective conception of probability and typical of Jeffreys' work.

something that involves no magnitude other than v, and if v is dimensional this can be satisfied only by a power of v. (Jeffreys 1967, 105)

In order to narrow the admissible probability density functions, Jeffreys appeals to a pair of criteria:

- The probability density function must be positive.
- For any $a \in (0, \infty)$, the probabilities of $(0, a)$ and (a, ∞) ought to be infinite.

While the first of these is axiomatic, the second represents an additional appeal to intuition. A truly "uninformative" prior, according to Jeffreys, would avoid assigning definitive weights to $(0, a)$ and (a, ∞) for any $a \in (0, \infty)$ since we don't actually know how the two intervals compare to one another. Even if we accept these criteria, they do not—contra Jeffreys' suggestion in the quote above—suffice to fix a probability density function of the form $Ax^n\ dx$. The (improper) probability density function $p_T : (0, \infty) \to [0, \infty)$ derived above provides a ready counterexample:

$$p_T(x) = \begin{cases} \frac{1}{x} & x \in (0, 1) \\ \frac{2}{2x-1} & \text{otherwise.} \end{cases}$$

It is easy to verify that p_T meets both criteria yet is not of the form $Ax^n\ dx$. Jeffreys' justification for considering only power transformations thus fails.

Indeed, the introduction of further intuitive requirements on "uninformative" priors only serves to worsen Jeffreys' situation. For all the intuitive appeal of Jeffreys' new constraint, it is again unclear why this restriction should not generalize beyond the bounds he provides, viz. intervals of the form $(0, a)$ and (a, ∞). Given complete ignorance about $x \in (0, \infty)$, is it not equally illicit to adopt determinate weights for two disjoint intervals $[a, b]$ and $[a', b']$? or, as Milne (1983) observes, to declare that the probability of $(0, a)$ is infinite while $[a, b]$ is finite? If we are truly ignorant, should we not avoid a determinate assignment to even singletons $\{a\}$ and $\{a'\}$? Just as with his initial appeal to invariance, taking Jeffreys new constraint seriously quickly threatens to deliver a set of restrictions which rules out all non-trivial measures. For all its intuitive appeal, Jeffreys' invariance program cannot succeed without serious modification.

Jeffreys' program was revived by Jaynes (1973) who not only extended the account to *Circles and Chords* but more importantly provided a new criteria for when invariance under a transformation was and was not required. Unlike Jeffreys, Jaynes relativized the choice of distribution not to the kind of parameter considered but rather to the problem posed. Jaynes held that problems themselves are resilient to certain transformations and so require solutions that are likewise invariant (Jaynes 1973; Rosenkrantz 1977). In the case of *Circles and Chords*, Bertrand does not bother, for example, to fix a particular orientation for the circle, nor a particular size, nor a particular spatial location. The problem itself is invariant under these transformations. Jaynes concludes that the desired likelihood must also be unaffected. That is, our reasoning in *Circles and Chords* ought also to be invariant under rotations, scale transformations, and translations. This inference from problem invariance to solution invariance has, as Jaynes observes, significant intuitive appeal. Anyone who proposes that the correct likelihood for *Circles and Chords* depends on the time of day or color of the circle is clearly to be ignored. The only novelty here is in recognizing that a number of mathematical transformations ought to be similarly irrelevant.

Specifying chords by the polar coordinates (d, θ) of their midpoint, Jaynes (1973) shows that requiring invariance under rotations, scale transformations, and translations suffices to uniquely determine a probability density function for *Circles and Chords*:

$$p(d, \theta) = \frac{1}{2\pi r d}$$

for $0 \leq d \leq r$ and $0 \leq \theta \leq 2\pi$. In fact, the requirement of translation invariance alone suffices to determine this probability density function (Jaynes 1973). Solving with respect to $p(d, \theta)$ identifies the correct likelihood as $\frac{1}{2}$ in agreement with Bertrand's first method, *Distance from Chord Midpoint to Origin*.

Jaynes' account readily extends to *Specific Volume and Density* and arguably to *The Mystery Cube Factory* as well. The key insight in both cases is that measuring physical quantities is often arbitrary in particular respects, and so both the problem and solution in such cases ought to be invariant along these same dimensions. The basics of this insight date to at least Keynes (1921, 50-51):

> The objective quality measured may not, strictly speaking, possess numerical quantitativeness, although it has the properties necessary for measurement by means of correlation with numbers. The values which it can assume may be capable of being ranged in an order, and it will sometimes happen that the series which is thus formed is continuous, so that a value can always be found whose order in the series is between any two selected values; but it does not follow from this that there is any meaning in the assertion that one value is twice another value. The relations of continuous order can exist between the terms of a series of values, without the relations of numerical quantitativeness necessarily existing also, and in such cases we can adopt a largely arbitrary measure of the successive terms, which yields results which may be satisfactory for many purposes, those, for instance, of mathematical physics, though not for those of probability.

A full response would nevertheless wait until Rosenkrantz (1977) and van Fraassen (1989).

Each of length, area, mass, and volume are equipped with a greatest lower bound or "natural zero" as well as an appropriate "addition" operation $\circ$. Taking length as an example, all values are at least as large as a single mark acting as both beginning and end. Any acceptable scale for lengths thus has a unique "zero" point. Lengths are similarly equipped with a salient "addition" operation $\circ$ by placing quantities end to end:

Following Krantz et al. (1971), these two characteristics suffice to fix real-valued measurement of a quantity up to choice of a unit. For all its formal sophistication, this point is readily apparent in the joint acceptability of measuring lengths in centimeters, inches, feet, and so forth; different scales which are all related by a constant factor α. More generally, length, area, mass, and volume are standard examples of quantities which admit of a ratio scale, a system of measurement m such that all and only similarity transformations,

$$T(m(x)) = \alpha m(x)$$

for $\alpha > 0$, produce a satisfactory measure.

Since all the physical quantities in both *Specific Volume and Density* and *The Mystery Cube Factory* are only fixed up to a choice of unit, the solution to each ought to be invariant under similarity transformations:

> *Similarity Invariance.* For any transformation $T(x) = \alpha x$ with $\alpha > 0$, the likelihood of a value in $I_{c,d}$ relative to a larger interval $I_{a,b}$ is equal to the likelihood of a value in $I_{T(c),T(d)}$ relative to the larger interval $I_{T(a),T(b)}$.

As the following argument shows, this requirement suffices to almost uniquely determine a solution:

> Consider $I_{a,b} = (0, \infty)$ and thus $I_{T(a),T(b)} = (0, \infty)$. Supposing that a continuous probability density function f exists for x and a continuous probability density function f_T exists for $T(x)$, the probability of an ϵ-interval dx is
>
> $$p(dx) = f(x)\, dx$$
>
> while the probability of an ϵ-interval $dT(x)$ is
>
> $$\begin{aligned} p(dT(x)) &= f_T(T(x))\, dT(x) \\ &= \alpha f_T(\alpha x)\, dx. \end{aligned}$$
>
> Since $x \in dx$ if and only if $T(x) \in d(T(x))$,
>
> $$f(x) = \alpha f_T(\alpha x).$$
>
> By *Similarity Invariance* over $I_{a,b} = I_{T(a),T(b)} = (0, \infty)$, f and f_T must be equivalent over every subinterval of $(0, \infty)$, and thus f and f_T must be the same function. It follows immediately that
>
> $$f(x) = \alpha f(\alpha x)$$
>
> for any $\alpha > 0$. This is uniquely satisfied by
>
> $$f(x) \propto \frac{1}{x}.$$

Requiring invariance under similarity transformations thus determines a prior over $(0, \infty)$ up to a choice of a constant.

Restricting this prior to a particular interval $I_{a,b} \subset (0, \infty)$ with $a > 0$, this choice of constant is determined by normality. The correct prior $p_{I_{a,b}}$ over an interval $I_{a,b}$ is thus the log-uniform density:

$$p_{I_{a,b}}(x) = \left(\frac{1}{\ln(b) - \ln(a)}\right)\frac{1}{x}.$$

The probability of a subinterval $I_{c,d}$ relative to a larger interval $I_{a,b} \subset (0, \infty)$ is then

$$p_{I_{a,b}}(I_{c,d}) = \frac{\ln(d) - \ln(c)}{\ln(b) - \ln(a)}.$$

Similarity Invariance thus fixes a unique class of improper probability density functions over $(0, \infty)$. Requiring normality over an interval $I_{a,b} \subset (0, \infty)$ with $a > 0$ then determines a unique function from this class.

Using the log-uniform density function with *Specific Volume and Density*, the probability of a specific volume in $[1, \frac{3}{2}]$ from a total space of $[1, 3]$ is

$$\begin{aligned} p_{[1,3]}([1, \tfrac{3}{2}]) &= \frac{\ln(\frac{3}{2}) - \ln(1)}{\ln(3) - \ln(1)} \\ &= \frac{\ln(3) - \ln(2)}{\ln(3)} \\ &\approx .369. \end{aligned}$$

Meanwhile, the probability of a density in $[\frac{2}{3}, 1]$ from a total space of $[\frac{1}{3}, 1]$ is

$$\begin{aligned} p_{[\frac{1}{3},1]}([\tfrac{2}{3}, 1]) &= \frac{\ln(1) - \ln(\frac{2}{3})}{\ln(1) - \ln(\frac{1}{3})} \\ &= \frac{\ln(3) - \ln(2)}{\ln(3)} \\ &\approx .369. \end{aligned}$$

The contradiction underlying *Specific Volume and Density* thus disappears when the log-uniform density is used. Despite their differences, Jaynes' approach to invariance resolves the paradox and identifies Jeffreys' proposal as the unique solution.

Applying the same procedure to *The Mystery Cube Factory* is complicated by the fact that $\frac{1}{x}$ diverges at 0 and leaves $\int_0^a \frac{1}{x}\,dx$ undefined for any $a \in (0, \infty)$. If we consider instead cubes with side-lengths in $I_{a,b} \subset (0, \infty)$ with $a > 0$, however, it is easy to verify that the log-uniform density again provides consistent probabilities after reformulation in terms of face-areas. More generally, for any $I_{a,b} \subset (0, \infty)$ with $a > 0$, adopting the log-uniform density provides invariant probabilities under any transformation

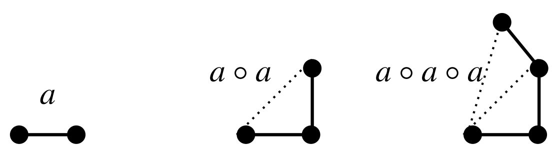
$$T(x) = \alpha x^{\beta}$$

for $\alpha > 0$ and $\beta \in \mathbb{R} - \{0\}$ since

$$\begin{aligned} p_{I_{T(a),T(b)}}(I_{T(c),T(d)}) &= \frac{\ln(\alpha d^{\beta}) - \ln(\alpha c^{\beta})}{\ln(\alpha b^{\beta}) - \ln(\alpha a^{\beta})} \\ &= \frac{\beta \ln(d) + \alpha - \beta \ln(c) - \alpha}{\beta \ln(b) + \alpha - \beta \ln(a) - \alpha} \\ &= \frac{\ln(d) - \ln(c)}{\ln(b) - \ln(a)} \\ &= p_{I_{a,b}}(I_{c,d}). \end{aligned}$$

While $p(x) \propto \frac{1}{x}$ is only invariant up to the choice of a constant for power transformations, adopting the normalized prior proportional to $\frac{1}{x}$ for a specific interval $I_{a,b} \subset (0, \infty)$ with $a > 0$ fixes the assigned probabilities under power transformations generally. While these probability density functions are not invariant simpliciter under power transformations, the probabilities they deliver relative to a specific interval $I_{a,b} \subset (0, \infty)$ with $a > 0$ are.

As Milne (1983) notes, the sudden concordance here between invariance under similarity transformations and invariance under power transformations has the air of happenstance. It seems a cosmic coincidence that the invariance requirement introduced by measurement also forces probabilistic invariance under the power transformations used by both *Specific Volume and Density* and *The Mystery Cube Factory*. Drawing on Ellis (1960), Milne proposes that this is due to an oversight in the foregoing analysis. While a "natural zero" and an intuitive "addition" operation suffice to fix measurement up to similarity, it does not follow that all measurement systems are thereby exhausted. Natural zeros and intuitive addition operations are suggestions rather than mandates. Though undoubtedly odd, it is also legitimate for example to ‘add’ lengths at 90° angles with the diagonal as the sum:



Setting a as 1 inch, Ellis shows that a length of x inches corresponds to a length of $\sqrt{x}$ diagonal inches or ‘dinches’ (Ellis 1960, 45). While not a similarity transformation of the usual measurement scheme, dinches are nevertheless a legitimate method of measuring length.

Specific Volume and Density and *The Mystery Cube Factory* thus require a prior which is invariant under not only similarity transformations but also alternative measurement schemes like Ellis’ dinches. Since dinches make use of the transformation $T(x) = x^{\frac{1}{2}}$, invariance is presumably required for power transformations in addition to similarity transformations. While requiring invariance over both power and similarity transformations leaves only trivial measures, Milne (1983) shows that weakening the invariance criteria to invariance up to a constant factor allows us to again establish $p(x) \propto \frac{1}{x}$ as the desired class of priors over $(0, \infty)$.

The invariance strategy thus claims a number of successes. Plausible solutions have been identified for not only Bertrand's *Circles and Chords* but also both *Specific Volume and Density* and *The Mystery Cube Factory*. Further, while the underlying measure-theoretic justification does not apply, simply adopting the log-uniform distribution in *Wine and Water* does again provide an invariant solution. As a result, Jeffreys' (1967) proposal to resolve Bertrand's paradox by identifying an appropriately invariant prior has dominated attempts to provide an objective solution, especially in light of Jaynes' (1973) suggested relativization to problems.

The Inadequacy of Invariance

Despite these successes, the invariance approach to Bertrand's paradox has ultimately proven inadequate. In general, invariance criteria need deliver neither a unique solution to an instance of Bertrand's paradox nor any solution at all. The resolutions cited above are happy accidents, cases when salient invariance criteria happen to align with a particular transformation from one parameterization to another. In general, these need not agree and so the invariance program does not provide a general resolution of Bertrand's paradox.

A simple modification of *Wine and Water* apparently due to Milne (1983) demonstrates the program's shortcomings. Milne asks after not just the proportion of wine to water and water to wine, but also the fraction of the total solution made up by wine:

> *Water and Wine - Wine/(Wine+Water)*
> A glass is presented which contains a mixture of water and wine. It is known only that the glass contains at least half as much wine as water and no more than an even mix, i.e., the fraction of wine in the glass $\frac{\text{wine}}{\text{wine+water}}$ is in $[\frac{1}{3}, \frac{1}{2}]$. How likely is it that the ratio of water to wine is in $[\frac{3}{7}, \frac{1}{2}]$?

While the log-uniform distribution provides consistent probabilities in both *Water and Wine - Wine/Water* and *Water and Wine - Wine/Water*, the shift to the fraction of wine relative to the whole reintroduces the paradox. This is, of course, because the transformation in question—$T(x) = \frac{x}{x+1}$—is not of the form αx^{β} for $\alpha > 0$ and $\beta \in \mathbb{R} - \{0\}$. Further, if we now add probabilistic invariance under $T(x) = \frac{x}{x+1}$ to our list of demands, the only solutions are trivial measures.

More generally, the invariance program never truly escaped the original criticism laid against Jeffreys. No non-trivial prior is invariant under all transformations, and so the invariance program must justify restricting our attention to only a particular set of transformations. Jaynes' proposal to relativize invariance criteria to problems only disguises this difficulty by emphasizing the particular transformation or transformations provided in the problem. Inconsistency is nevertheless guaranteed under some bijective transformation, and so the paradox can always be reintroduced by specifying an additional transformation. In the end, neither Jeffreys nor Jaynes nor their later proponents provide a compelling reason for restricting the transformations we consider. As a result, proving that this or that set of transformations suffices to eliminate all or almost all priors is entirely irrelevant. No non-trivial measure is invariant under all acceptable transformations.

5.1.3 The Standard Analysis of Bertrand's Paradox

Discussion of Bertrand's paradox has largely focused on resolving particular instances rather than characterizing the underlying phenomenon. One of the few commentators to diverge from this trend is Keynes (1921, 52):

> In general, if x and $f(x)$ are both continuous variables, varying always in the same or in the opposite sense, and x must lie between a and b, then the probability that x lies between c and d, where $a < c < d < b$, seems to be $\frac{d-c}{b-a}$, and the probability that $f(x)$ lies between $f(c)$ and $f(d)$ seems to be $\frac{f(d)-f(c)}{f(b)-f(a)}$. These expressions, which represent the probabilities of necessarily concordant conclusions, are not, as they ought to be, equal.

Nearly a century later, Bangu (2009, 31) provides a nearly identical analysis:

> One begins with a variable x uniformly distributed in an interval $[a, b]$ and then one considers a scaling transformation θ such that $x' = \theta(x)$. For a fixed value c such that $a \leq c \leq b$, two questions (or 'problems', as Bertrand calls them) are formulated:
>
> Q1: What is $p_1 = P(x \in [c, b]$, if x is random in $[a, b])$?
>
> Q2: What is $p_2 = P(x' \in [c', b']$, if x' is random in $[a', b'])$?
>
> These two questions are said to be 'identical', so they should receive the same answer. However, if we calculate the probabilities using [the principle of indifference] (and the standard Lebesgue measure of the intervals), we notice that $p_1 = |b - c|/|b - a|$ is different from $p_2 = |b' - c'|/|b' - a'|$, for many transformations θ, and values of a, b and c.

Though less explicit, Jeffreys (1967), Rosenkrantz (1977, 1981), Gillies (2000), Mikkelson (2004), and Howson and Urbach (2006) all sketch similar accounts.

Let $I_{a,b}$ denote an interval from $\mathbb{R}$ of the form $[a, b], [a, b), (a, b]$, or (a, b) with $a < b$. The common core for these accounts can be summarized as:

STANDARD ANALYSIS OF BERTRAND'S PARADOX:

Every instance of Bertrand's paradox invokes a space of possibilities Ω and an event $E \subseteq \Omega$ together with

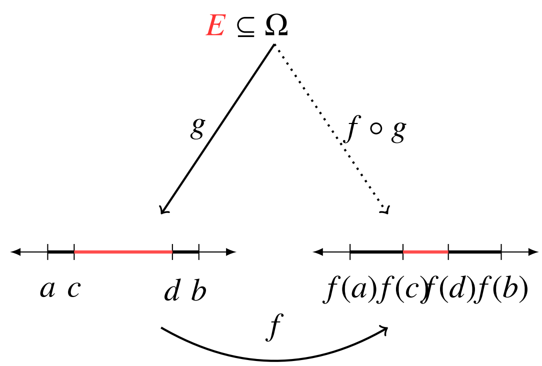
(i) a bijection $g : \Omega \to I_{a,b}$ such that $g[E] = I_{c,d}$ and

(ii) a bijection $f : I_{a,b} \to I_{f(a),f(b)}$ (or $I_{f(b),f(a)}$) such that $f[I_{c,d}] = I_{f(c),f(d)}$ (or $I_{f(d),f(c)}$)

where

(iii) $\frac{|d-c|}{|b-a|} \neq \frac{|f(d)-f(c)|}{|f(b)-f(a)|}$.

Further, in so far as this structure clearly obtains, we likewise have an intuitive instance of Bertrand's paradox.



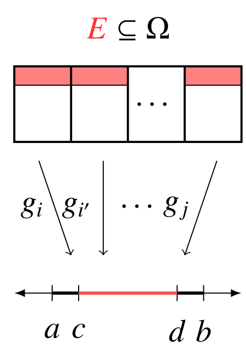
On this analysis, the paradox is induced by representing the space of possibilities[8] Ω in terms of two intervals $I_{a,b}$ and $I_{f(a),f(b)}$ wherein some single event E—represented by the two subintervals $I_{c,d}$ and $I_{f(c),f(d)}$—receives two different "sizes", namely $\frac{|d-c|}{|b-a|}$ and $\frac{|f(d)-f(c)|}{|f(b)-f(a)|}$.

Applying the analysis to the *The Mystery Cube Factory*, the space of possibilities Ω is the set of cubes which can be produced by the factory, viz. cubes with side-length ≤ 2 cm and face-area ≤ 4 cm^2. The event E in question is then the desired subset of these cubes, viz. cubes with side-length ≤ 1 cm and face-area ≤ 1 cm^2. Note that both sets are well defined. Problems arise, however, when this space and event are viewed under two intuitive bijections, g and $f \circ g$. The first of these bijections, g, maps the set of possible cubes to their side-lengths in $[0, 2]$. The second bijection, $f \circ g$, maps the set of possible cubes to their face-areas in $[0, 4]$. To complete the diagram, note that side-lengths and face-areas are themselves related by an intuitive bijection $f(x) = x^2$. Finally, the size of E relative to all possible cubes varies depending on which of g and $f \circ g$ is used, satisfying condition (iii) of the standard analysis.

[8]The existence of a well-defined set of possibilities (or sample space) is sometimes missed by commentators, e.g., Pettigrew (2016). Nevertheless, this is a key part of the paradox. A situation wherein different likelihoods are ascribed over different sets of possibilities is unremarkable; the contradictory solutions are troubling precisely because the underlying space and event are fixed.

As Rowbottom and Shackel (2010) have also observed, *Circles and Chords* does not conform to the standard analysis.[9] Of his three methods, only Bertrand's last—*Area of Midpoint Regions*—appears to supply a bijection between the set of all possible chords and another space. This other space, however, is $\mathbb{R}^2$ not $\mathbb{R}$, and Bertrand uses relative area rather than relative interval length to generate the associated likelihood. To make matters still more complicated, there is only the appearance of a bijection here because, while any point in a circle besides the center is the midpoint of a unique chord, the center of a circle is the midpoint for uncountably many chords. Even if we emphasize appearance and overlook this failure to construct a bijection, Bertrand's third method still breaks with the standard analysis by using $\mathbb{R}^2$ rather than $\mathbb{R}$.

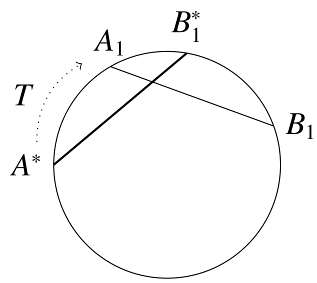
Bertrand's first and second methods retain the use of intervals on $\mathbb{R}$ but drop all pretense of a bijection with the set of all chords. Bertrand's first method, *Distance from Origin to Chord Midpoint*, appears to partition the set of all chords Ω and then supplies a collection of bijections from the parts of this partition which not only all map into the same interval but which also all send their share of chords longer than $\sqrt{3}r$ to the same subinterval.



Two chords $\overline{A_1B_1}$ and $\overline{A_2B_2}$ share a part in the apparent partition if and only if they are parallel and their midpoints lie along the same radius. As Bertrand observes, the distance of a chord's midpoint from the center of the circle defines a natural bijection between any single part in this partition and the interval $[0, r]$. Finally, chords from any part are longer than $\sqrt{3}r$ if and only if they fall into $[0, \frac{r}{2})$ under their associated bijection. The likelihood of a chord longer than $\sqrt{3}r$ with respect to any part is thus $\frac{1}{2}$. Since this result holds for every part, the overall likelihood of a chord longer than $\sqrt{3}r$ is also $\frac{1}{2}$.

Bertrand's second method, *Angle from Tangent*, has a parallel structure. Fix a point A^* on the circle. An apparent partition of the set of all chords can now be had by placing two chords $\overline{A_1B_1}$ and $\overline{A_2B_2}$ in the same part if and only if for a particular rotation T, there exists B_1^* and B_2^* such that $T(\overline{A^*B_1^*}) = \overline{A_1B_1}$ and $T(\overline{A^*B_2^*}) = \overline{A_2B_2}$.

[9]While it is possible to construct a superficially similar situation which does fit, Bertrand clearly makes use of premises which deviate from those of the standard analysis. Since paradoxes (or instances of paradoxes) are individuated in part by their premises, constructing clever bijections cannot save the standard analysis here. It is important not that (i)-(iii) could be satisfied for Ω and E but rather that (i)-(iii) are the premises which lead to paradox.



Each part is then generated by taking the chords with A^* as an endpoint and rotating by some fixed amount. Since the set of chords with A^* as an endpoint can be put into bijection with the interval $[0°, 180°)$ using the degree θ from the tangent at A^*, every part in this partition can likewise be put into bijection with $[0°, 180°)$ by the same process. Finally, chords from any part are longer than $\sqrt{3}r$ if and only if they fall into $(60°, 120°)$ under their associated bijection. The likelihood of a chord longer than $\sqrt{3}r$ with respect to any part is thus $\frac{1}{3}$. Since this result holds for every part, the overall likelihood of a chord longer than $\sqrt{3}r$ is also $\frac{1}{3}$.

Despite their intuitive appeal, both arguments follow Bertrand's third method in playing fast and loose with chords which run through the center of the circle. The "partitions" defined for both the first and second methods are no such thing (Shackel 2007; Rowbottom 2013). Fix a chord $\overline{AB}$ which intersects the center of the circle. For the construction used in the first method, note that there are two distinct radii which both intersect the midpoint of $\overline{AB}$ and are perpendicular to $\overline{AB}$, viz. one in each direction. It follows immediately that chords perpendicular to each radius share a "part" with $\overline{AB}$ but not with each other. It follows immediately that the apparent partition underlying the first method is ill-defined.[10]

For the second method, the failure of the apparent partition can be demonstrated by taking the "part" consisting of all chords with A as an endpoint and rotating 180°. In most cases, this rotation behaves as intended and produces a new chord. Rotating $\overline{AB}$ by 180°, however, produces $\overline{AB}$. The apparent partition underlying Bertrand's second method thus assigns $\overline{AB}$ to both the original "part" and the "part" generated by a 180° rotation. It follows immediately that the apparent partition underlying the second method is also ill-defined. Close examination of both Bertrand's first and second methods thus reveals a systematic double counting of chords which intersect the center of the circle.

Circles and Chords thus greatly complicates the status of the standard analysis. Taking all three methods seriously suggests that the standard analysis ought to be generalized along two distinct dimensions. First, Bertrand's use of relative area in his third method suggests that the paradox ought to extend beyond the relative length of intervals. Second, Bertrand's use of partitions in his first and second methods suggests that the paradox does not require explicit bijections for the entire space of possibilities. At the same time, there is good reason to object to all three methods, potentially undercutting the case for these generalizations altogether.

[10] As Shackel (2007) observes, the first method can be repaired by dropping reference to a radius and putting two chords in a part if and only if they are parallel. This is not, however, Bertrand's method.

5.1.4 Revised Analysis of Bertrand's Paradox

In light of the standard analysis, Bertrand's paradox is strongly associated with continuous intervals, relative interval lengths, and the mathematical theory of probability. Despite this long association, none of these are an essential feature of the paradox. Rather, the core of the paradox is a conflict between the intuition that bijections ought to preserve relative sizes and a host of particular measures which do not. The two generalizations suggested by *Circles and Chords* are well taken.

Infinite collections have long been associated with paradox. Zeno's paradoxes are a famous case from antiquity but difficulties continued throughout the middle ages and well into the 20th century. Galilei (1638) provides a classic example with the intuitive attractiveness of both (1) and (2):

(1) Natural numbers are more numerous than squares of natural numbers.

(2) There are just as many squares of natural numbers as there are natural numbers.

The initial plausibility of (1) stems from the *part-whole* intuition, the conviction that proper parts are smaller than the whole. Since

$$\{0, 1, 4, 9, 16, \ldots\} \subset \{0, 1, 2, 3, 4, \ldots\}$$

it ought to be that the natural numbers are strictly larger than set of all squares of natural numbers. (2) meanwhile is supported by a similarly appealing *bijection invariance* intuition, the conviction that bijections between sets preserve size. Since squaring natural numbers

$$\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \ldots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ 0 & 1 & 4 & 9 & 16 & \ldots \end{array}$$

clearly produces a bijection between the natural numbers and their squares, these sets ought to be the same size. As Mancosu (2009) argues at length, the proper response to this situation is to recognize that both *part-whole* and *bijection invariance* encode consistent intuitions about size but that the move to infinite sets requires a choice between them.

Bertrand's paradox is a similar situation only for the relative size of sets. Suppose we wish to know how the relative size of a set E compares to the relative size of a set E' where $E, E' \subseteq \Omega$. If Ω is not itself equipped with a salient notion of relative size, it is natural to represent Ω in terms of a set which is so equipped and thereby obtain our desired comparison. Bertrand's paradox appears when two or more of these representations contradict one another.

Revised Analysis of Bertrand's Paradox:

Every instance of Bertrand's paradox invokes a space of possibilities Ω and events $E, E' \subseteq \Omega$ together with

(i') a set U_1 and notion of relative size $\precsim_1$ such that either

  (a) there exists a bijection $g_1 : \Omega \to U_1$ or

  (b) there exists a partition P of Ω such that, for any choice of $P_i \in P$, there is a corresponding bijection $g_1^i : P_i \to U_1$, and for any i, j, both

$$g_1^i[E \cap P_i] = g_1^j[E \cap P_j]$$

$$g_1^i[E' \cap P_i] = g_1^j[E' \cap P_j]$$

(ii') a set U_2 and notion of relative size $\precsim_2$ such that either

  (a) there exists a bijection $g_2 : \Omega \to U_2$ or

  (b) there exists a partition P of Ω such that, for any choice of $P_i \in P$, there is a corresponding bijection $g_2^i : P_i \to U_2$, and for any i, j, both

$$g_2^i[E \cap P_i] = g_2^j[E \cap P_j]$$

$$g_2^i[E' \cap P_i] = g_2^j[E' \cap P_j]$$

where

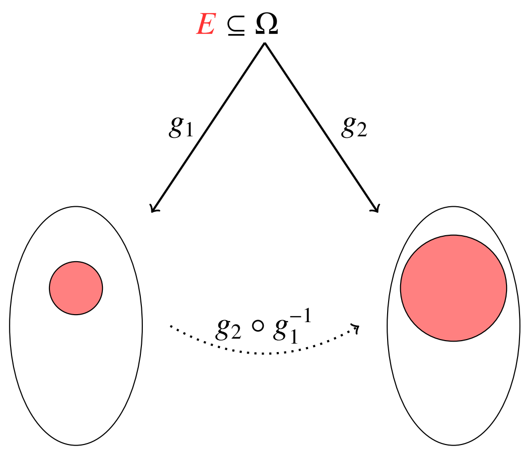
(iii') $\precsim_1$ and $\precsim_2$ do not agree on the relative sizes of E and E' (under g_1 and g_2 respectively).

Further, in so far as this structure clearly obtains, we likewise have an intuitive instance of Bertrand's paradox.

$$\begin{tikzcd}
& E, E' \subseteq \Omega \arrow[dl, "g_1"'] \arrow[dr, "g_2"] & \\
g_1[E] \precsim_1 g_1[E'] \arrow[rr, dotted, bend right, "g_2 \circ g_1^{-1}"] & & g_2[E] \not\precsim_2 g_2[E']
\end{tikzcd}$$

More precisely, Bertrand's paradox is the result of two representations $\langle U_1, \precsim_1 \rangle$ and $\langle U_2, \precsim_2 \rangle$ which engender conflicting relative size intuitions over two sets E and E' drawn from a common superset Ω. A set U and notion of relative size $\precsim$ may qualify as a representation either by means of a bijection $g : \Omega \to U$ or by means of a collection of uniform bijections from the parts in some partition of Ω to U. A conflict between representations can be constructed whenever relative size is not preserved under bijection, viz. when it is not the case that $E \precsim E'$ if and only if $g[E] \precsim g[E']$ for any bijection $g : \Omega \to \Omega'$.

In the case where U_1 and U_2 are equipped not just with comparative size relations $\precsim_1$ and $\precsim_2$ but quantitative measures μ_1 and μ_2 over a shared scale, explicit reference to a second set E' can be dropped and (iii') collapsed into a disagreement over the relative size of E, namely $\mu_1(g_1[E]) \neq \mu_2(g_2[E])$. Under these circumstances, we obtain a clear generalization of the diagram for the standard analysis:



While the standard instances of Bertrand's paradox all use quantitative measures of relative size, purely comparative variations of, for example, *The Mystery Cube Factory* are easy to construct.

All of the standard instances of Bertrand's paradox can be explained on the revised analysis. *The Mystery Cube Factory* presents a space of possibilities Ω containing all cubes of side-length $\leq$ 2cm (and thus also face-areas $\leq$ 4cm^2) and then asks after the relative size of a particular subset E, the cubes with side-length $\leq$ 1 cm and face-areas $\leq$ 2 cm^2. The paradox arises when these sets of cubes are distilled into just lengths ($g_1[E]$ and U_1) and areas ($g_2[E]$ and U_2) respectively, representations which possess intuitive measures of relative size which vary under bijection. The underlying comparative nature of the disagreement can be seen by comparing E with its complement E^c, the set of cubes with a side-length of $1-2$ cm and face-area of $1-4$ cm^2. In the side-length representation, $g_1[E^c] \precsim_1 g_1[E]$ but in the face-area representation $g_2[E^c] \not\precsim_2 g_2[E]$. The pattern repeats verbatim with *Water and Wine*. Again, a well-defined set of water and wine mixtures is presented and a particular subset, $\frac{water}{wine}$ ratios within $[1, 2]$ and $\frac{wine}{water}$ ratios within $[\frac{1}{2}, 1]$, is distinguished. Representing this single set in terms of only one ratio or the other again suggests relative interval length as an intuitive measure of size for each. By construction, the result is a pair of inconsistent sizes despite clear bijections.

Unlike *The Mystery Cube Factory* and *Wine and Water*, *Circles and Chords* only appears to conform to the revised analysis. *Circles and Chords* considers all chords on a circle of radius r (the total space Ω) and then asks after all these chords longer than $\sqrt{3}r$ (the subset E). The first method in *Circles and Chords* opts for the appearance of case (b), a partition of Ω wherein every part P_i admits of representation by the interval $[0, r]$, and $E \cap P_i$ is always identified with the subinterval $[0, \frac{r}{2})$. As Bertrand illustrates, this uniform representation across parts intuitively entails a matching likelihood for the union of these parts, licensing the move from a likelihood of $\frac{1}{2}$ for E in every part to a likelihood of $\frac{1}{2}$ for E full stop. The second method likewise provides the appearance of satisfying case (b), providing an apparent partition and collection of bijections which not only map every part P_i to $[0, 180°)$ but also $E \cap P_i$ to $[60°, 120°]$. The third method in *Circles and Chords* breaks with the previous two and supplies an apparent bijection from the set of all chords on a circle of radius r to points within the same circle. This latter space comes equipped with a natural measure of relative size in the form of relative area. Regardless of its actual shortcomings, *Circles and Chords* thus appears to satisfy the revised analysis and thereby generates an apparent instance of the paradox.[11]

Just as the paradox of infinity presented by Galilei (1638) forces a choice between bijection invariance for sizes and the part-whole intuition, so too does Bertrand's paradox force a choice between bijection invariance for relative sizes and various intuitive notions thereof. Within the current literature, only Shackel (2007), Norton (2008), and Gyenis and Rédei (2014) hint at such a conception of Bertrand's paradox. Norton (2008) is, however, led astray by his emphasis on descriptions while both Shackel (2007) and Gyenis and Rédei (2014) restrict themselves to the probability formalism. None recognize that Bertrand's paradox is not inherently tied to likelihood at all. Properly understood, Bertrand's paradox concerns all and only the conflict between the intuition that relative size is preserved under bijection and measures of relative size which are not so preserved.

The adequacy of the revised analysis requires not only a satisfactory treatment for accepted instances of Bertrand's paradox but also that further instances of the revised analysis give rise to further cases of the paradox. To motivate this latter claim, I present two new instances of Bertrand's paradox which deviate significantly from canonical cases along with a general schema for generating further instances. For $I_{c,d} \subseteq I_{a,b} \subseteq \mathbb{R}$, the natural measure of relative size is relative interval length:

$$\mu_{\mathbb{R}}(I_{c,d}, I_{a,b}) = \frac{d-c}{b-a}.$$

These same intuitions, however, readily provide a natural measure of relative size over

[11] As the phrasing here suggests, I am inclined to reclassify *Circles and Chords* as only an apparent instance of Bertrand's paradox. This is owed in large part to the availability of instances which don't depend on the formal errors underlying Bertrand's methods. Anyone who wishes to retain *Circles and Chords* as a proper instance of the paradox, however, need only weaken the revised analysis so that it stipulates the appearance of the presented structure.

rational intervals as well. For $I_{c,d} \subseteq I_{a,b} \subseteq \mathbb{Q}$, define

$$\mu_{\mathbb{Q}}(I_{c,d}, I_{a,b}) = \frac{d-c}{b-a}.$$

Since $\mu_{\mathbb{Q}}$—like $\mu_{\mathbb{R}}$—is not preserved under bijection, the revised analysis holds that an instance of Bertrand's paradox is available.

A simple modification of *Wine and Water* suffices. Consider a glass containing a mixture of wine and water where the two liquids were mixed in some rational ratio[12] with at least half as much wine as water and no more than equal parts wine and water, i.e., $\frac{\text{water}}{\text{wine}} \in [1,2]$ and $\frac{\text{wine}}{\text{water}} \in [\frac{1}{2}, 1]$. Focusing on the proportion of water to wine, the likelihood that the mix is a rational number in $[1, \frac{4}{3}]$—$\mu_{\mathbb{Q}}([1, \frac{4}{3}], [1,2])$—is once again intuitively $\frac{1}{3}$. Considering instead ratios of wine to water, all rational numbers in $[\frac{1}{2}, 1]$ are possible. The likelihood that the wine to water ratio is in $[\frac{3}{4}, 1]$—$\mu_{\mathbb{Q}}([\frac{3}{4}, 1], [\frac{1}{2}, 1])$—is once again intuitively $\frac{1}{2}$. Since water to wine ratios and wine to water ratios are reciprocals, the two intuitive solutions of $\frac{1}{3}$ and $\frac{1}{2}$ directly contradict one another. Bertrand's paradox extends to intervals of rational numbers and thus beyond the probability formalism.

The use of a dense[13] space is similarly inessential. For an infinite collection Y of natural numbers and $X \subseteq Y$, one intuitive notion of relative size is given by a minor generalization of natural density:

$$\mu_{\mathbb{N}}(X, Y) = \lim_{n \to \infty} \frac{|X_{\leq n}|}{|Y_{\leq n}|}$$

for $n \in \mathbb{N}$ where

$$X_{\leq n} = \{x \in X \mid x \leq n\}$$
$$Y_{\leq n} = \{y \in Y \mid y \leq n\}.$$

It is this notion of size which leads us to say, for instance, that even numbers make up half the natural numbers—$\mu_{\mathbb{N}}(\text{Even}, \mathbb{N}) = \frac{1}{2}$—while multiples of four make up one quarter of the naturals and one half of all evens—$\mu_{\mathbb{N}}(\text{MultipleOf4}, \mathbb{N}) = \frac{1}{4}$ and $\mu_{\mathbb{N}}(\text{MultipleOf4}, \text{Even}) = \frac{1}{2}$.

Since $\mu_{\mathbb{N}}$ ascribes different sizes to the set of even numbers and the set of multiples of four relative to $\mathbb{N}$, $\mu_{\mathbb{N}}$ does not preserve relative size under bijection. This can be immediately leveraged into an intuitive instance of Bertrand's paradox:

> *Tupperware Factory.* A factory produces tupperware with a natural number n printed on the lid and another m printed on the bottom. Further, every natural number appears exactly once in each position. How many of the tupperware have an even number on their lid? How many of the tupperware have a multiple

[12]For those tempted to label such a case outlandish, note that it is no less wild to think that any possible real value was available in the original case. Appearances aside, a glass of water is composed of only a finite number of H_2O molecules.

[13]An (partial or total) ordering $\langle X, \leq \rangle$ is dense if and only if for any $x_1, x_2 \in X$ such that $x_1 < x_2$, there exists $x \in X$ satisfying $x_1 < x < x_2$.

of four on their bottom?

The intuitive responses (à la $\mu_{\mathbb{N}}$) are that evens make up exactly $\frac{1}{2}$ of all possible lid values, and so the first likelihood is $\frac{1}{2}$. It is similarly plain that multiples of four make up exactly $\frac{1}{4}$ of all possible bottom values, and so the second likelihood is $\frac{1}{4}$. Bertrand's paradox now appears when it is revealed that the factory pairs lids and bottoms in the somewhat odd fashion described by f:

$$f(x) = \begin{cases} 2x & \text{if } x = 2n \text{ for some } n \in \mathbb{N} \\ n + 1 + (n+1)\,\mathrm{Div}\,4 & \text{if } x = 2n+1 \text{ for some } n \in \mathbb{N}. \end{cases}$$

The function f pairs all even numbered lids with bottoms bearing a multiple of four and shuffles all odd numbered lids to the remaining bottoms:

$$\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & \dots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ 0 & 1 & 4 & 2 & 8 & 3 & \dots \end{array}$$

As a result, the set of tupperware with an even numbered lid is the set of tupperware with a bottom bearing a multiple of four, and so the intuitive values of $\frac{1}{2}$ and $\frac{1}{4}$ contradict one another.

As promised, this particular puzzle easily generalizes into a schema for further instances of Bertrand's paradox whenever relative sizes are not preserved under bijection:

> *Generalized Tupperware Factory.* A factory produces tupperware with elements from U_1 printed on the lid and elements from U_2 printed on the bottom. All elements from U_1 and U_2 are used exactly once in this process. Finally, these markings are related by a particular function f so that, should the lid and bottom of a tupperware accidentally be separated, the appropriate match can be identified. How do the tupperware with an element from $g_1[E]$ on their lid compare (more, less, equal, incomparable) to those with an element from $g_1[E']$ on their lid? How do the tupperware with an element from $g_2[E]$ on their bottom compare (more, less, equal, incomparable) to those with an element from $g_2[E']$ on their bottom?

If the answer for $g_1[E]$ and $g_1[E']$ differs from that for $g_2[E]$ and $g_2[E']$, f is unveiled as matching lids with elements from $g_1[E]$ to bottoms marked with elements from $g_2[E]$ as well as lids with elements from $g_1[E']$ to bottoms marked with elements from $g_2[E']$. *Generalized Tupperware Factory* thus forces a confrontation between these relative size judgments and the clear bijection between tupperware lids and tupperware bottoms.

While the revised analysis greatly expands the scope of Bertrand's paradox, not all intuitive measures of relative size are susceptible. The most obvious example of a relative measure which is immune to the paradox is counting elements in finite sets. For a finite set Ω and $E \subseteq \Omega$, define

$$\mu_{\text{Fin}}(E, \Omega) = \frac{|E|}{|\Omega|}.$$

Unlike previous examples, this measure is fixed under any bijection $g : \Omega \to \Omega'$ and thus satisfies the bijection invariance intuition.

Proposition 5.1

For any finite Ω and $A \subseteq \Omega$, if $g : \Omega \to \Omega'$ is a bijection, then

$$\mu_{Fin}(A, \Omega) = \mu_{Fin}(g[A], g[\Omega]).$$

Supposing that the counting measure μ_{Fin} is the only acceptable metric for relative size over finite sets, Bertrand's paradox cannot then arise in a finite setting. An infinite set of possibilities is an essential component of the paradox, and so Bertrand's paradox can be properly accounted a paradox of infinity.

Finally, given an infinite set of possibilities and a few particular sets in $\mathcal{F}$, all non-trivial measures are susceptible to Bertrand's paradox:

Proposition 5.2

Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\mu : \mathcal{F} \to [0, \infty]$ a measure. Suppose that

(i) *Ω is infinite,*

(ii) *$B \in \mathcal{F}$ is an infinite set with $0 < \mu(B) < \infty$, and*

(iii) *there exists $A, A', A'_1, A'_2 \in \mathcal{F}$ such that $|A| = |A'| = |A'_1| = |A'_2|$, A, A' partition B, and A'_1, A'_2 partition A'.*

Then, there exists a bijection $g : \Omega \to \Omega$ such that

$$\mu(A) \neq \mu(g[A]).$$

As a matter of pure mathematics, this measure-instability can be blocked by either removing enough propositions from $\mathcal{F}$ or by sufficiently restricting the available bijections g. The philosophical basis for either strategy with respect to a theory of confirmation is suspect in the extreme. As a result, the measure-theoretic framework—and thus contemporary probability theory—is incompatible with an objective resolution to Bertrand's paradox.

5.2 First-Order Confirmation

The previous chapter laid the foundations for a theory of confirmation based on the classical equipossibility account of probability. These foundations are summarized in the following pair of principles:

(0) The collection of all canonical-domain models is an exhaustive collection of atomic possibilities.

(1') If $\Omega_{\mathcal{L}}$ is an exhaustive collection of atomic possibilities, then every possibility in the subcollection $\Omega_{\mathcal{L},\Sigma}$ is an equal possibility relative to $\Omega_{\mathcal{L}}$ and Σ.

Unlike these first two, the last component of the finite confirmation account was left almost entirely unchanged from Laplace's presentation of the equipossibility account.

(2') For a finite, non-empty set of equal possibilities, the degree of confirmation for a sentence φ is the number of equal possibilities which make φ true divided by the total number of equal possibilities.

The restriction to only finite sets of equal possibilities here is severe, and so an extension of (2') to infinite sets is desired.

The primary barrier to extending (2') is Bertrand's paradox. While early formulations of the paradox only undercut intuitive measures—relative interval length and relative area in particular—the revised analysis defended in the previous section shows that an instance of Bertrand's paradox can be constructed whenever a metric for relative size varies under bijection. This criteria straightforwardly rules out not just a number of intuitive measures for particular infinite sets but any non-trivial measure. Confirmation theory demands a notion of relative size which is more general than that offered by contemporary probability or even measure theory.

Echoing Jeffreys' influential discussion, the "uninformative distribution" for a parameter known only to lie in $(0, \infty)$ ought to be invariant under power transformations. Having no information about a parameter x drawn from $(0, \infty)$ entails a corresponding lack of information about the parameter x^2 also drawn from $(0, \infty)$. There is no reason, however, to stop with only power transformations. If one truly has no information about a parameter x drawn from $(0, \infty)$, then—beyond what is encoded in g itself—one has no information about $g(x)$ for any transformation g. Even if we take the lesson of the previous chapter to heart and make explicit that each value in $(0, \infty)$ represents an atomic possibility, any bijection g still induces a parameter $g(x)$ and a second set of atomic possibilities $g[(0, \infty)]$. The "uninformative distribution" demands invariance under not just power transformations or similarity transformations but any bijection.

Jeffreys' desire for particular incomparabilities—though he never uses the word—was also well conceived but overly restricted. If a parameter is known only to lie in $(0, \infty)$ with each value atomic, then one ought not endorse a definitive relation between $(0, a)$ and (a, ∞) for any $a \in (0, \infty)$ on pain of contradiction. There is, however, no reason to stop with just these sets. If one truly knows only that x lies in $(0, \infty)$ with each value atomic, then a definitive relation between any pair of disjoint intervals $I_{a,b}$ and $I_{a',b'}$ is similarly illicit. The uninformative distribution ought to really, truly be uninformative. In so far as Jeffreys' initial discussion is compelling, it ought to convince us that no probability density function (proper or improper) supplies an adequate "uninformative" prior and thus no adequate solution to Bertrand's paradox exists in the probability formalism.

As stringent as these restrictions appear, the finite confirmation account developed in the previous chapter as well as the minimalist confirmation theories $\lesssim_{\Sigma}^{\text{PL}}$ and $\lesssim_{\Sigma}^{\text{FOL}}$ show that confirmation is not thereby reduced to triviality. The current section refines the proposed theory of confirmation by providing a comparative extension over infinite collections. The resulting confirmation relation is strictly stronger than the confirmation theory induced by first-order logic, validates almost all the axioms of comparative probability, and exhibits the bijection invariance required in order to avoid Bertrand's paradox. Further, it is a maximal solution in the sense that any refinement either violates the cardinality ordering, violates monotonicity, or is susceptible to Bertrand's paradox.

5.2.1 Extending the Equipossibility Account to Infinite Collections

The equipossibiliity account ranks sentences by taking the number of favorable possibilities divided by the total number. This scheme fails immediately in an infinite context since neither the notion of number nor the division operation dictate a particular extension to infinite sets. This difficulty can be partially alleviated by noting that the total number of possibilities serves only to introduce an absolute scale; comparative relationships between sentences are determined by the number of favorable possibilities alone. Suppressing the total number of possibilities thus provides us the comparative maxim that greater confirmation corresponds to more favorable possibilities.

Applying this comparative maxim to an infinite collection of atomic possibilities $\Omega_{\mathcal{L},\Sigma}$ delivers a large number of intuitive confirmation relations. First, the confirmation ranking again ranges between two extremes. At one end are sentences which have no favorable possibilities in $\Omega_{\mathcal{L}}$, claims inconsistent with the supplied evidence Σ. On the other side are sentences for which every possibility in $\Omega_{\mathcal{L}}$ is a favorable case, claims which are entailed by the supplied evidence Σ. In between these two extremes are sentences which correspond to all the remaining subcollections of $\Omega_{\mathcal{L},\Sigma}$.

The simplest of these intermediary sentences are those which correspond to finite sets. Consistency with the rankings adopted for finite $\Omega_{\mathcal{L},\Sigma}$ dictates that these are ranked by number of atomic possibilities. Any two sentences which correspond to the same number of atomic possibilities are thus to be accounted equally likely while any sentence which corresponds to exactly two atomic possibilities is to accounted less likely than any sentence which corresponds to three atomic possibilities and so forth. Having recognized a significant number of comparative confirmation relations between finite sets, it is tempting to immediately impose a mirror structure on their complements, the co-finite collections. Any sentence which corresponds to all atomic possibilities except two, for example, ought to be more likely than any sentence which corresponds to all atomic possibilities except three since on the other end of the ranking three atomic possibilities are more likely than two. The substantial difficulty in extending the equipossibility account to infinite sets is thus only in what to do with sentences which correspond to collections which are both infinite and co-infinite.

The core of the proposed extension is to leverage a relativized form of (2') and to declare incomparability for whatever remains.

Definition 5.1 Let Ω be a collection and A, B subcollections. Define

$A \prec^{\dagger}_{\Omega} B$ if and only if there exists a one-to-one $g : A - B \to B - A$ but not vice-versa.

$A \sim^{\dagger}_{\Omega} B$ if and only if $A - B$ and $B - A$ are finite with $|A - B| = |B - A|$.

$A \precsim^{\dagger}_{\Omega} B$ if and only if $A \prec^{\dagger}_{\Omega} B$ or $A \sim^{\dagger}_{\Omega} B$.

For a first-order language L, collection of canonical-domain models $\Omega_{\mathcal{L}}$ for L, and evidence set $\Sigma \subseteq L$, the comparative confirmation relation $\precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}}$ may be interpreted as a relation over L by mapping sentences to the collection of canonical-domain models in $\Omega_{\mathcal{L},\Sigma}$ which make them true, i.e., $\varphi \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} \psi$ if and only if $[\![\varphi]\!]_{\Omega_{\mathcal{L},\Sigma}} \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} [\![\psi]\!]_{\Omega_{\mathcal{L},\Sigma}}$. In addition, when Ω is clear from context, $\precsim^{\dagger}_{\Omega}$, $\prec^{\dagger}_{\Omega}$, and $\sim^{\dagger}_{\Omega}$ will be abbreviated to $\precsim^{\dagger}$, $\prec^{\dagger}$, and $\sim^{\dagger}$.

In the event that Ω is a set, the proposed comparative confirmation relation $\precsim^{\dagger}_{\Omega}$ has a straightforward relationship to cardinalities. The definition of the strict relation, for example, collapses into

$$A \prec^{\dagger}_{\Omega} B \text{ if and only if } |A - B| < |B - A|$$

More generally, any two sets A and B fall into one of the following four cases:

- Case (i): $A - B$ finite and $B - A$ finite.
  $A \precsim^{\dagger}_{\Omega} B$ if and only if $|A - B| \leq |B - A|$.
- Case (ii): $A - B$ finite and $B - A$ infinite.
  $A \prec^{\dagger}_{\Omega} B$.
- Case (iii): $B - A$ finite and $A - B$ infinite.
  $B \prec^{\dagger}_{\Omega} A$.
- Case (iv): $A - B$ infinite and $B - A$ infinite.
  $A \precsim^{\dagger}_{\Omega} B$ if and only if $|A - B| < |B - A|$.

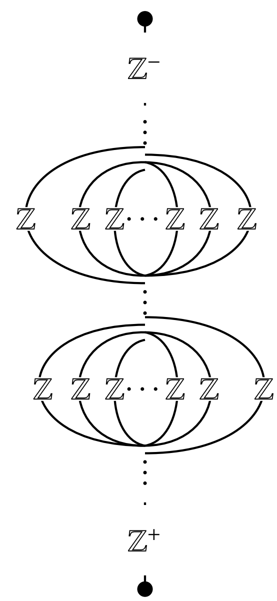
Case (i) is a simple relativization of the method of equipossibilities, endorsing the finite ordering when the difference in the favorable possibilities between two sets is finite. Cases

(ii) and (iii) extend this to the mixed finite-infinite case by stipulating that an infinity of additional favorable possibiliities is larger than any finite number. The most complicated component of the proposed extension is case (iv). Comparisons between infinite $A - B$ and infinite $B - A$ are endorsed only when there is a difference of cardinality; otherwise, the two sets are accounted incomparable to one another.

Taking $\Omega = \mathbb{N}$ as an example, we have the intuitive ranking over finite and co-finite sets while infinite sets which are also co-infinite exhibit a much more complicated structure. Taking the set E of even numbers and the set O of odd numbers as an example, E and O are incomparable to one another as are any sets reachable within a finite number of additions or subtractions:

$$
\begin{array}{rcl}
 & \begin{array}{l} [\Omega] = \mathbb{N} \\ \curlyvee^{\dagger} \\ [\Omega - 1] = \mathbb{N} - \{0\} \sim^{\dagger} \mathbb{N} - \{1\} \sim^{\dagger} \ldots \\ \curlyvee^{\dagger} \\ [\Omega - 2] = \mathbb{N} - \{0,1\} \sim^{\dagger} \mathbb{N} - \{0,2\} \sim^{\dagger} \ldots \\ \curlyvee^{\dagger} \\ \vdots \end{array} & \\
\begin{array}{r} \cdots \sim^{\dagger} E \cup \{3\} \sim^{\dagger} E \cup \{1\} = [E+1] \\ \curlyvee^{\dagger} \\ \cdots \sim^{\dagger} E \cup \{1\} - \{2\} \sim^{\dagger} E = [E] \\ \curlyvee^{\dagger} \\ \cdots \sim^{\dagger} E - \{4\} \sim^{\dagger} E - \{2\} = [E-1] \end{array} & \begin{array}{c} \iddots \quad \ddots \\ \cdots \\ \ddots \quad \iddots \end{array} & \begin{array}{l} [O+1] \sim^{\dagger} O \cup \{2\} \sim^{\dagger} O \cup \{4\} \sim^{\dagger} \ldots \\ \curlyvee^{\dagger} \\ [O] = O \sim^{\dagger} O \cup \{2\} - \{1\} \sim^{\dagger} \ldots \\ \curlyvee^{\dagger} \\ [O-1] = O - \{1\} \sim^{\dagger} O - \{3\} \sim^{\dagger} \ldots \end{array} \\
 & \begin{array}{l} \vdots \\ \curlyvee^{\dagger} \\ [2] = \{0,1\} \sim^{\dagger} \{0,2\} \sim^{\dagger} \{1,2\} \sim^{\dagger} \ldots \\ \curlyvee^{\dagger} \\ [1] = \{0\} \sim^{\dagger} \{1\} \sim^{\dagger} \{2\} \sim^{\dagger} \ldots \\ \curlyvee^{\dagger} \\ [0] = \emptyset \end{array} &
\end{array}
$$

Abstracting further and more fully leveraging the indexing in terms of $\mathbb{Z}$, the defined ordering provides a total order at either extreme (isomorphic to $\mathbb{Z}^{+}$ and $\mathbb{Z}^{-}$ respectively) with an infinite number of incomparable $\mathbb{Z}$-chains in between:



While focusing on only a collection of disjoint infinite and co-infinite sets—e.g., the evens and odds—delivers a discrete ball of incomparabilities as pictured above, overlapping infinite and co-infinite sets will further complicate the diagram by producing overlapping balls.

Despite this complexity, $\precsim^{\dagger}$ satisfies all of de Finetti's (1951) axioms for intuitive probability with the sole exception of comparability:

Proposition 5.3

Let Ω be given. Then for any choice of $A, B, C \in \mathcal{P}(\Omega)$, $\precsim^{\dagger}_{\Omega}$ satisfies

C0. Nontriviality
$\emptyset \prec \Omega$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3a. Nonnegativity
$\bot \precsim A$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

Failures of comparability are moreover restricted to the infinite case. With a finite collection of atomic possibilities, $\precsim^{\dagger}$ satisfies all of de Finetti's axioms:

Proposition 5.4

Let Ω be finite. Then for any choice of $A, B \in \mathcal{P}(\Omega)$, $\precsim^{\dagger}_{\Omega}$ satisfies

C1b. Comparability
$A \precsim B$ or $B \precsim A$.

In addition, $\precsim^{\dagger}$ satisfies associativity as well as a number of intuitive principles not endorsed by the probability formalism:

Proposition 5.5

Let Ω be given. Then for any choice of $A, B, C, C' \in \mathcal{P}(\Omega)$, $\precsim^{\dagger}_{\Omega}$ satisfies

FE. Finite Equivalence
If A, B finite and $|A| = |B|$, then $A \sim B$.

CFE. Co-Finite Equivalence
If A, B co-finite and $|A^c| = |B^c|$, then $A \sim B$.

R. Regularity
If $A \neq \emptyset$, then $\emptyset \prec A$.

PW. Part-Whole
If $A \subset B$, then $A \prec B$.

FD. Finite Difference
If C, C' finite with $|C| < |C'|$ and $A \cap C = A \cap C' = \emptyset$, then $A \cup C \prec A \cup C'$.

SC. Strong Cardinality
If $|A| < |B|$, then $A \prec B$.

BI. Bijection Invariance
For any bijection $g : \Omega \to \Omega'$, $A \precsim B$ if and only if $g[A] \precsim g[B]$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, and $G_4 \sim F_3 \cup E_3$ with matching subscripts disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

5.2.2 Confirmation for First-Order Languages

Revising (2') to make exclusive use of the $\precsim^{\dagger}$ relation completes the proposed account of first-order confirmation.

Confirmation for First-Order Languages

Let L be a first-order language and $\Sigma \subseteq L$ a consistent[14] set of evidence.

(0) The collection of all canonical-domain models is an exhaustive collection of atomic possibilities.

(1') If $\Omega_{\mathcal{L}}$ is an exhaustive collection of atomic possibilities, then every possibility in the subcollection $\Omega_{\mathcal{L},\Sigma}$ is an equal possibility relative to $\Omega_{\mathcal{L}}$ and Σ.

(2") For a collection of equal possibilities $\Omega_{\mathcal{L},\Sigma}$ with subcollections A and B, A is at least as confirmed as B if and only if $A \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B$.

While the comparative confirmation relation $\precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}}$ properly ranges over $\mathcal{F}^{+}_{\mathcal{L},\Sigma}$, it may be unambiguously lifted to $A, B \in \mathcal{F}^{+}_{\mathcal{L}}$ by setting $A \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B$ if and only if $A \cap \Omega_{\mathcal{L},\Sigma} \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B \cap \Omega_{\mathcal{L},\Sigma}$.

Over arbitrary collections of atomic possibilities, the first-order confirmation account is a strict extension of the relation induced by first-order logic. $\precsim^{\text{FOL}}_{\Sigma}$ and $\precsim^{\dagger}$ both satisfy C01a23b4A as well as regularity, part-whole, and bijection invariance. $\precsim^{\dagger}$, however, also endorses a number of additional comparisons based on the number of atomic possibilities which compose an event or correspond to a given sentence. In finite cases, this delivers a total preorder; in infinite ones, it suffices for principles like finite equivalence, co-finite equivalence, finite difference, and strong cardinality.

Given only a set of evidence Σ in a formal language L, however, the comparative confirmation relation $\precsim^{\dagger}$ over the collection of all canonical-domain models which make Σ true and the comparative confirmation relation $\precsim^{\text{FOL}}_{\Sigma}$ induced by first-order logic agree over L.

Proposition 5.6

Let L be a first-order language and $\Sigma \subseteq L$ a consistent theory. Taking all canonical-domain models for L and any extension of L as $\Omega_{\mathcal{L}}$, for any $\varphi, \psi \in L$,

$$\varphi \precsim^{\dagger}_{\Sigma} \psi \text{ if and only if } \varphi \precsim^{FOL}_{\Sigma} \psi.$$

[14] As earlier, this is consistency with respect to $\Omega_{\mathcal{L}}$ if $\Omega_{\mathcal{L}}$ is supplied. If $\Omega_{\mathcal{L}}$ is not supplied, this is first-order consistency or, equivalently, consistency with respect to the collection of all canonical-domain models.

The proposed account of confirmation is thus an extension of logic in two distinct senses. First, $\varphi \precsim_{\Sigma}^{\text{FOL}} \psi$ entails $\varphi \precsim^{\dagger} \psi$ regardless of $\Omega_{\mathcal{L}}$. Second, $\precsim_{\Sigma}^{\text{FOL}}$ and $\precsim^{\dagger}$ agree over L in the extreme case where only a set of evidence is provided. An unrestricted space of atomic possibilities thus reduces first-order confirmation to first-order logic while meaningful restrictions produce a strictly stronger relation.

5.2.3 A Maximal Resolution of Bertrand's Paradox

The comparative confirmation relation $\precsim^{\dagger}$ suffices to provide a unique comparative solution to any instance of Bertrand's paradox. In the *The Mystery Cube Factory* for example, it is true that a cube with side-length less than 1 cm is no more confirmed than a cube with side-length greater than 1 cm. Contrary to the intuitive assignment of likelihood $\frac{1}{2}$ to each, however, this does not entail that these claims are equally confirmed. Despite its intuitive appeal, $\varphi \sim^{\dagger} \psi$ does not follow from $\varphi \not\prec^{\dagger} \psi$ and $\psi \not\prec^{\dagger} \varphi$ in infinite contexts. Working through the definition above, *The Mystery Cube Factory* is asking after claims which are, according to $\precsim^{\dagger}$, incomparable to one another. This pattern repeats for all canonical instances of the paradox; in every case, incomparable events are being forced into a salient total order.

Incomparability is not a particularly desirable result. It is thus natural to ask if some of the incomparabilities in $\precsim^{\dagger}$ could be broken one way or the other. This is not possible without collapsing large portions of the current ranking. So long as we require bijection invariance, $\precsim^{\dagger}$ uniquely maximizes size distinctions:

Proposition 5.7

Let Ω be a set and $\precsim$ a binary relation on $\mathcal{P}(\Omega)$ satisfying

WC. Weak Cardinality
If $|A| < |B|$, then $B \not\prec A$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

PI. Permutation Invariance
For any permutation $\pi : \Omega \to \Omega$, $A \precsim B$ if and only if $\pi[A] \precsim \pi[B]$.

Then, $A \prec B$ only if $A \prec^{\dagger} B$.

Supposing (WC), (C4), and (PI), the only means of eliminating incomparabilities is thus to sacrifice relative size distinctions between sets. Given the alternatives, it is the part-whole principle which naturally suffers:

Corollary 5.1

Let Ω be given and let $\precsim$ be a strict extension of $\precsim^\dagger$. Then, at least one of the following is false:

WC. Weak Cardinality
If $|A| < |B|$, then $B \not\prec A$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

PI. Permutation Invariance
For any permutation $\pi : \Omega \to \Omega$, $A \precsim B$ if and only if $\pi[A] \precsim \pi[B]$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

PW. Part-Whole
If $A \subset B$, then $A \prec B$.

These sacrifices, moreover, are neither gradual nor local. Supposing (C4), (WC), (PI), and (C2), breaking only a single incomparability in $\precsim^\dagger$ entails the collapse of all distinctions between finite sets and all distinctions between co-finite sets:

Corollary 5.2

Let Ω be given and let $\precsim$ be a strict extension of $\precsim^\dagger$ satisfying:

WC. Weak Cardinality
If $|A| < |B|$, then $B \not\prec A$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

PI. Permutation Invariance
For any permutation $\pi : \Omega \to \Omega$, $A \precsim B$ if and only if $\pi[A] \precsim \pi[B]$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

Then for any $A^*, B^* \in \mathcal{P}(\Omega)$ such that $A^* \precsim B^*$ but $A^* \not\precsim^\dagger B^*$ and $B^* \not\precsim^\dagger A^*$,

(i) $\emptyset \sim C$ *for any* $C \in \mathcal{P}(\Omega)$ *with* $|C| \leq |A^* - B^*|$ *and*

(ii) $\Omega \sim C$ *for any* $C \in \mathcal{P}(\Omega)$ *with* $|C^c| \leq |A^* - B^*|$.

Any attempt to eliminate incomparabilities in the proposed ranking entails violation of an even more fundamental principle. The cost of even a single additional comparison is giving up either (PI), (WC), (C4), (C2), or the obvious truth that a claim with some favorable cases is more confirmed than a claim which contradicts our evidence. None of these is an acceptable price in a theory of confirmation, and so every incomparability in $\precsim^\dagger$ is necessary. The comparative confirmation relation $\precsim^\dagger$ is a maximal resolution of Bertrand's paradox.

5.2.4 Absolute Rankings and Degrees of Confirmation

Historically, accounts of confirmation—most notably Carnap (1962)—have been formulated as absolute rankings assigning "degrees of confirmation". Since $\precsim^\dagger$ satisfies C01b23a4A, the work of Chapter two entails that the comparative confirmation relation $\precsim^\dagger$ over any particular $\Omega_{\mathcal{L},\Sigma}$ can be represented as a likelihood assignment into the (minimal representing) likelihood structure induced by $\precsim^\dagger$, $\langle \mathbb{P}_{\precsim^\dagger}, \circ_{\precsim^\dagger} \rangle$. It is easy to verify, however, that no single likelihood structure suffices for every choice of $\Omega_{\mathcal{L},\Sigma}$ merely on cardinality grounds. As a result, the proposed account of confirmation is not categorical, and thus there is no fixed set which can be identified as all and only the degrees of confirmation. While this breaks with both contemporary formalizations of rational credence and historical accounts of confirmation, it is far from unexpected. The weight of a single possibility intuitively ought to shrink as the set of all atomic possibilities grows first through finite and then infinite sets of greater and greater cardinality.[15]

Talk of degrees of confirmation is nevertheless intuitively appealing, particularly if we wish to make comparisons between the confirmation of A in Ω_A and the confirmation of B in Ω_B. The first two subsections below expand on the relationship between the comparative confirmation relation $\precsim^\dagger$ and absolute rankings, showing that the formal structure of $\precsim^\dagger$ precludes unique representation in any likelihood structure outside of the finite case. Despite the lack of both categoricity and unique representation, the third and final subsection presents a minimalist extension of the comparative confirmation relation $\precsim^\dagger$ to cross-set confirmation comparisons. Taking equivalence classes under this cross-set relation as "degrees of confirmation" then provides a rigorous basis for degree of confirmation talk.

[15]This is sometimes presented as an objection to regularity intuitions, e.g., in Hájek (2010). This objection is only persuasive if we retain the view that circumscribing a set of all degrees of confirmation is possible. Considering a single possibility relative to larger and larger total sets should convince proponents of regularity that demanding a fixed set of all confirmation values is a mistake. As the set of atomic possibilities increases so too does the set of distinct confirmation values.

Absolute Rankings in the Finite Case

Over a finite set of models $\Omega_{\mathcal{L},\Sigma}$, the comparative confirmation relation $\precsim^{\dagger}$ is naturally identified with the absolute ranking $\mu : \mathcal{F}^{+}_{\mathcal{L},\Sigma} \to \mathbb{Q} \cap [0,1]$ described by (2'),

$$\mu(E) = \frac{|E|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

Proposition 5.8

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then $\precsim^{\dagger}$ on $\mathcal{F}^{+}_{\mathcal{L},\Sigma}$ is uniquely represented in $\langle \mathbb{Q} \cap [0,1], \leq, + \rangle$ by

$$\mu(E) = \frac{|E|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

This correspondence also readily extends to the algebra $\mathcal{F}^{+}_{\mathcal{L}}$.

Corollary 5.3

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then $\precsim^{\dagger}$ on $\mathcal{F}^{+}_{\mathcal{L}}$ is uniquely represented in $\langle \mathbb{Q} \cap [0,1], \leq, + \rangle$ by

$$\mu^{+}(E) = \frac{|E \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

$\precsim^{\dagger}$ is thus a generalization of the absolute ranking proposed in the equipossibility account.

The existence of a unique representing assignment into a fixed likelihood structure naturally supports a degree of realism about the values in this structure. In the case of $\precsim^{\dagger}$ over finite sets, this impression is further reinforced by the fact that these values are preserved under uniform model space expansion of any finite size.

Definition 5.2 For a first-order language L, extension L^{+} of L, and $\mathcal{M}^{+}$ a canonical-domain model for L^{+} or an extension thereof,[16] $\mathcal{M}^{+}|^{L^{+}}_{L}$ is the canonical-domain model $\mathcal{M}$ obtained by eliminating all symbols in the signature of L^{+} and not in the signature of L from $\mathcal{M}^{+}$.

[16] Recall that the models in a model space for a formal language L need not be L-models; rather, they must assign truth values over L. It is thus legitimate to include L^{+}-models in a model space for L so long as L^{+} is an extension of L.

Definition 5.3 For $\Omega_{\mathcal{L}}$ a model space for L and $\Omega_{\mathcal{L}^+}$ a model space for L^+ with $L \subseteq L^+$, $\Omega_{\mathcal{L}^+}$ is an *n-uniform model space expansion* of $\Omega_{\mathcal{L}}$ for $n \in \mathbb{N}^+$ if and only if there exists a partition $\{P_i, \dots\}$ of $\Omega_{\mathcal{L}^+}$ such that for any i, j:

$$|P_i| = |P_j| = n;$$

$$\text{For any } \mathcal{M} \in \Omega_{\mathcal{L}} \text{ there exists a unique } P_i \text{ such that } \mathcal{M}^+ \in P_i \Rightarrow \mathcal{M}_i^+|_L^{L^+} = \mathcal{M};$$

$$\text{For any } P_i \text{ there exists a unique } \mathcal{M} \in \Omega_{\mathcal{L}} \text{ such that } \mathcal{M}^+ \in P_i \Rightarrow \mathcal{M}^+|_L^{L^+} = \mathcal{M}.$$

An n-uniform model space expansion thus extends every model in $\Omega_{\mathcal{L}}$ to exactly n models in $\Omega_{\mathcal{L}^+}$. Because models are always extended, there is also no difficulty in extending n-uniform model space expansion to cases including a evidential set Σ.

Proposition 5.9

Let $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+}\rangle$ be an n-uniform extension of $\mathcal{L} = \langle L, \Omega_{\mathcal{L}}\rangle$ and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then, there exists a partition $\{P_i, \dots\}$ of $\Omega_{\mathcal{L}^+,\Sigma}$ such that for any i, j:

$$|P_i| = |P_j| = n;$$

$$\textit{For any } \mathcal{M} \in \Omega_{\mathcal{L},\Sigma} \textit{ there exists a unique } P_i \textit{ such that } \mathcal{M}^+ \in P_i \Rightarrow \mathcal{M}_i^+|_L^{L^+} = \mathcal{M};$$

$$\textit{For any } P_i \textit{ there exists a unique } \mathcal{M} \in \Omega_{\mathcal{L},\Sigma} \textit{ such that } \mathcal{M}^+ \in P_i \Rightarrow \mathcal{M}^+|_L^{L^+} = \mathcal{M}.$$

Finally, uniformly decomposing each possibility into a set of n possibilities leaves the value assigned to "events" fixed.

Proposition 5.10

Let $\Omega_{\mathcal{L}^+}$ be an n-uniform model space expansion of $\Omega_{\mathcal{L}}$ and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then for any $E \in \mathcal{F}^+_{\mathcal{L},\Sigma}$,

$$\mu(E) = \mu^+(E^+)$$

where $E^+ = \{\mathcal{M}^+ \in \Omega_{\mathcal{L}^+,\Sigma} \ : \ \mathcal{M}^+|_L^{L^+} \in E\}$ and the functions $\mu : \mathcal{F}^+_{\mathcal{L},\Sigma} \to \mathbb{Q} \cap [0,1]$, $\mu^+ : \mathcal{F}^+_{\mathcal{L}^+,\Sigma} \to \mathbb{Q} \cap [0,1]$ are the unique representing assignments of $\precsim^\dagger$ into $\langle \mathbb{Q} \cap [0,1], \leq, +\rangle$.

The finite case thus admits of not only unique assignments relative to a fixed likelihood structure but also fixed values under decomposition of the possibility space.[17] The natural

[17]This fact engenders a degree of ambivalence towards the atomicity or non-atomicity of possibilities; so long as decompositions are uniform and the total set of possibilities is finite, it does not matter for our choice of μ. This finite-case ambivalence seems a likely culprit for the classical equipossibility account's failure to recognize the importance of atomic possibilities.

impression is that degrees of confirmation are (or at least are measured or represented by) values from $\langle \mathbb{Q} \cap [0, 1], + \rangle$.

Absolute Rankings in the Infinite Case

This impression rapidly unravels with infinite sets of atomic possibilities. Not only are assignments into $\langle \mathbb{Q} \cap [0, 1], + \rangle$ no longer adequate to represent $\precsim^{\dagger}$ but the formal properties of the comparative confirmation relation $\precsim^{\dagger}$ undercut the possibility of unique assignments altogether. The formal difficulties here are two-fold.

First, infinite sets of atomic possibilities introduce incomparabilities. Taking assignments into a scale with symmetrically-structured incomparable values guarantees a multiplicity of viable assignments since permuting which sets receive which incomparable values produces further viable assignments. In the following example, for instance, permuting the values assigned to the intermediate sets $\{\mathcal{M}_1\}$ and $\{\mathcal{M}_2\}$ produces a second viable assignment.

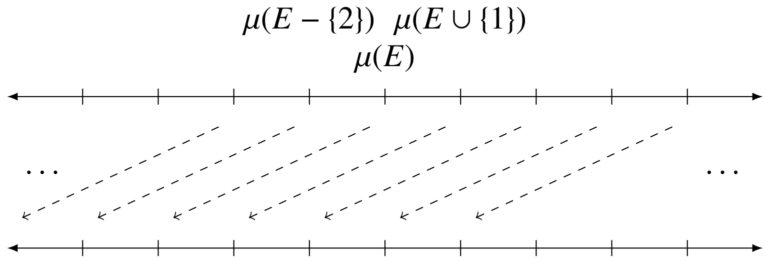
$$\begin{aligned} \mu(\{\mathcal{M}_1, \mathcal{M}_2\}) &= 1 \\ \mu(\{\mathcal{M}_1\}) &= \frac{1}{2}_a \\ \mu(\{\mathcal{M}_2\}) &= \frac{1}{2}_b \\ \mu(\emptyset) &= 0 \end{aligned}$$

$$\begin{tikzcd} & 1 & \\ \frac{1}{2}_a \arrow[ur, phantom, "\prec"] & & \frac{1}{2}_b \arrow[ul, phantom, "\prec"] \\ & 0 \arrow[ul, phantom, "\prec"] \arrow[ur, phantom, "\prec"] & \end{tikzcd}$$

Since $\precsim^{\dagger}$ produces an infinite number of symmetrically-structured incomparable sequences, any adequate likelihood structure for $\precsim^{\dagger}$ admits of an infinite number of viable assignments.

This situation is further exacerbated by the existence of $\mathbb{Z}$-chains—collections of sets whose ordering is isomorphic to $\langle \mathbb{Z}, \leq \rangle$—within the $\precsim^{\dagger}$ preorder. In both the finite and co-finite case, values may be fixed by reference to $\dot{0}$ and $\dot{1}$ respectively. With the intermediate infinite and co-infinite sets, however, there is no salient, finitely-reachable value off which to index an assignment. Since the structure of a $\mathbb{Z}$-chain is preserved under any finite shift in either direction, a single viable assignment again guarantees a multiplicity of such assignments.

Consider, for example, the set of even numbers E if all natural numbers $\mathbb{N}$ are a possibility.



Shifting the values assigned to the sequence $\ldots, E-\{2,4\}, E-\{2\}, E, E\cup\{1\}, E\cup\{1,3\}, \ldots$ by a finite amount in either direction produces a viable assignment over the sequence; propagating this shift throughout the entire assignment gives another assignment which represents $\precsim^{\dagger}$.

In contrast with the finite case, the formal structure of the comparative confirmation ranking $\precsim^{\dagger}$ over infinite sets of atomic possibilities guarantees an infinite number of representing assignments whenever a likelihood structure suffices for a single assignment. This is not to say that the proposed theory does not admit of absolute rankings nor that—in the terminology of §2.1—cross-set comparisons are never available. The comparative confirmation preorder $\precsim^{\dagger}$ over an infinite set simply does not admit of unique representations. As an immediate consequence, there is little reason to regard a salient likelihood structure as definitive of degrees of confirmation; the realism typical of, for example, probabilism is untenable in the proposed account. Confirmation is most directly and intuitively encoded by the $\precsim^{\dagger}$ relation.

Cross-Set Confirmation Comparisons

The most immediate consequence of losing unique representation in some particular likelihood structure $\langle \mathbb{P}, \circ \rangle$ is the loss of the cross-set confirmation comparisons induced by this representation. As §2.1 noted, two absolute rankings μ_A and μ_B over different spaces $\mathcal{P}(\Omega_A)$ and $\mathcal{P}(\Omega_B)$ which share a scale $\langle \mathbb{P}, \circ \rangle$ naturally extend to an absolute ranking μ^+ over $\Big(\mathcal{P}(\Omega_A) \times \{\Omega_A\}\Big) \cup \Big(\mathcal{P}(\Omega_B) \times \{\Omega_B\}\Big)$ by

$$\mu(\langle C, \Omega_C \rangle) = \begin{cases} \mu_A(C) & \text{if } \Omega_C = \Omega_A \\ \mu_B(C) & \text{if } \Omega_C = \Omega_B. \end{cases}$$

In this way, the likelihood of A in Ω_A may be compared to the likelihood of B in Ω_B even when Ω_A and Ω_B are distinct. With the canonical representation of $\precsim^{\dagger}$ over finite sets in terms of functions into $\langle \mathbb{Q} \cap [0,1], + \rangle$, for example, this operation ranks all pairs $\langle A, \Omega_A \rangle$ where Ω_A is a finite set of canonical-domain models and $A \subseteq \Omega_A$. All events E assigned the value $\frac{1}{2}$ are then accounted equally confirmed even if the underlying sets of atomic models differ. With the loss of unique representation in any likelihood structure, a new approach is required for cross-set confirmation comparisons over arbitrary collections of atomic models.

The remainder of this section focuses on extending the comparative confirmation relation $\precsim^\dagger$ to pairs $\langle A, \Omega_A\rangle$. The strategy here is to build up the stock of positive cross-set comparisons by means of intuitively plausible conditions without contradicting $\precsim^\dagger$ itself. A first, minimalist principle for extending the comparative confirmation relation $\precsim^\dagger$ to pairs $\langle A, \Omega_A\rangle$ where $A \subseteq \Omega_A$ is partwise comparison in the collection of all atomic models Ω.

P. *Partwise Comparison*
If both $A \precsim^\dagger_\Omega B$ and $\Omega_B \precsim^\dagger_\Omega \Omega_A$, then $\langle A, \Omega_A\rangle \precsim \langle B, \Omega_B\rangle$.

Definition 5.4 Let Ω be the collection of all atomic possibilities, Ω_A and Ω_B non-empty subcollections, and A, B subcollections of Ω_A and Ω_B, respectively. Define

$$\langle A, \Omega_A\rangle \precsim^P \langle B, \Omega_B\rangle \text{ if and only if } \langle A, \Omega_A\rangle \precsim \langle B, \Omega_B\rangle \text{ is entailed by (P).}$$

Informally, if A is at least as small as B relative to all possibilities and Ω_B is at least as small as Ω_A relative to all possibilities, then A in Ω_A must be smaller than B in Ω_B since it does worse in both relevant components.

While straightforward, $\precsim^P$ does not suffice for a number of expected cross-set confirmation comparisons. This is most obvious in the failure to fix either a minimal or maximal degree of confirmation:

$$\langle \emptyset, \{\mathcal{M}_1\}\rangle \not\precsim^P \langle \emptyset, \{\mathcal{M}_1, \mathcal{M}_2\}\rangle$$
$$\langle \{\mathcal{M}_1\}, \{\mathcal{M}_1\}\rangle \not\precsim^P \langle \{\mathcal{M}_1, \mathcal{M}_2\}, \{\mathcal{M}_1, \mathcal{M}_2\}\rangle.$$

In general, n-uniform model space expansions also generate distinct degrees of confirmation under $\precsim^P$, e.g.,

$$\langle \{\mathcal{M}_1\}, \{\mathcal{M}_1, \mathcal{M}_2\}\rangle \not\precsim^P \langle \{\mathcal{M}_{1.1}, \mathcal{M}_{1.2}\}, \{\mathcal{M}_{1.1}, \mathcal{M}_{1.2}, \mathcal{M}_{2.1}, \mathcal{M}_{2.2}\}\rangle$$

even if $\mathcal{M}_{i.j}$ is an extension of $\mathcal{M}_i$. Assuming we wish to recover at least those cross-set comparative confirmation judgments induced by representing $\precsim^\dagger$ over finite sets in $\langle \mathbb{Q} \cap [0,1], +\rangle$, a strict extension of $\precsim^P$ is required.

Stipulating preservation under uniform model space expansions suffices to make up the shortfall in the finite case.

U. *Finite Uniform Expansion*
If Ω_A^+ is an n-uniform model space expansion of Ω_A and Ω_A is finite, then $\langle A, \Omega_A\rangle \sim \langle A^+, \Omega_A^+\rangle$ where A^+ is the collection of models from Ω_A^+ which extend a model from A.

Uniform, finite model space expansions in other words preserve degrees of confirmation.

In isolation, (P) and (U) need not deliver a transitive relation. For example, if $\mathcal{M}_{i.j}$ is an extension of $\mathcal{M}_i$, then

$$\langle \{\mathcal{M}_1\}, \{\mathcal{M}_1, \mathcal{M}_2\}\rangle \precsim \langle \{\mathcal{M}_{1.1}, \mathcal{M}_{1.2}\}, \{\mathcal{M}_{1.1}, \mathcal{M}_{1.2}, \mathcal{M}_{2.1}, \mathcal{M}_{2.2}\}\rangle$$

follows from (U). From (P), we also have that

$$\langle\{\mathcal{M}_{1.1}, \mathcal{M}_{1.2}\}, \{\mathcal{M}_{1.1}, \mathcal{M}_{1.2}, \mathcal{M}_{2.1}, \mathcal{M}_{2.2}\}\rangle \precsim \langle\{\mathcal{M}'_1, \mathcal{M}'_2\}, \{\mathcal{M}'_1, \mathcal{M}'_2, \mathcal{M}'_3\}\rangle.$$

Nevertheless, neither (P) nor (U) require that

$$\langle\{\mathcal{M}_1\}, \{\mathcal{M}_1, \mathcal{M}_2\}\rangle \precsim \langle\{\mathcal{M}'_1, \mathcal{M}'_2\}, \{\mathcal{M}'_1, \mathcal{M}'_2, \mathcal{M}'_3\}\rangle.$$

At minimum, the cross-set comparisons delivered by (P) and (U) must thus be closed so that resulting relation is transitive.

Definition 5.5 Let Ω_A be a non-empty collection of canonical-domain models, and A a subcollection of Ω_A. Let $\precsim^{PU}$ denote the transitive closure of those comparisons required by (P) $\wedge$ (U) over pairs $\langle A, \Omega_A\rangle$.

$\precsim^{PU}$ is the cross-set comparative confirmation relation induced by representation in $\langle \mathbb{Q} \cap [0,1], +\rangle$.

Proposition 5.11

Let Ω_A and Ω_B be finite but non-empty sets of canonical-domain models with $A \subseteq \Omega_A$ and $B \subseteq \Omega_B$. Then, $\langle A, \Omega_A\rangle \precsim^{PU} \langle B, \Omega_B\rangle$ if and only if $\mu_{\Omega_A}(A) \leq \mu_{\Omega_B}(B)$ where $\mu_{\Omega_A} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ and $\mu_{\Omega_B} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ are the unique representations of $\precsim^{\dagger}_{\Omega_A}$ and $\precsim^{\dagger}_{\Omega_B}$ in $\langle \mathbb{Q} \cap [0,1], +\rangle$.

(P), (U), and transitivity thus suffice for all canonical comparisons over finite sets of atomic possibilities.

Further, it is easy to verify that the cross-set comparison relation $\precsim^{PU}$ agrees with $\precsim^{\dagger}$ within collections of arbitrary size.

Proposition 5.12

Let Ω_A be a collection of canonical-domain models and A, A' subcollections of Ω_A. Then,

$$\langle A, \Omega_A\rangle \precsim^{PU} \langle A', \Omega_A\rangle \text{ if and only if } A \precsim^{\dagger}_{\Omega_A} A'.$$

$\precsim^{PU}$ is thus a consistent extension of $\precsim^{\dagger}$ to pairs $\langle A, \Omega_A\rangle$.

The cross-set comparison relation $\precsim^{PU}$ nevertheless fails to capture some important intuitions. Most saliently, it fails to fix both a unique minimal and maximal degree of confirmation. For example,

$$\langle\emptyset, \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \dots\}\rangle \not\precsim^{PU} \langle\emptyset, \{\mathcal{M}_1\}\rangle$$

and

$$\langle\{\mathcal{M}_1,\mathcal{M}_2,\mathcal{M}_3,\dots\},\{\mathcal{M}_1,\mathcal{M}_2,\mathcal{M}_3,\dots\}\rangle \not\sim^{PU} \langle\{\mathcal{M}_1\},\{\mathcal{M}_1\}\rangle.$$

Neither of these are acceptable, and so a strict extension of $\precsim^{PU}$ is desired.

Given the existing endorsement of (U), one natural strategy is to generalize (U) by lifting either the restriction to finite sets or the restriction to partitions of finite size. Unfortunately, neither generalization is tenable. Lifting the restriction to partitions of finite size is straightforwardly inconsistent with $\precsim^{\dagger}$. Consider, for example, that

$$\langle\{\mathcal{M}_1\},\{\mathcal{M}_1,\mathcal{M}_2\}\rangle \sim^{\dagger} \langle\{\mathcal{M}_2\},\{\mathcal{M}_1,\mathcal{M}_2\}\rangle$$

but

$$\langle\{\mathcal{M}'_1,\mathcal{M}'_3,\mathcal{M}'_5,\dots\},\{\mathcal{M}'_1,\mathcal{M}'_2,\mathcal{M}'_3,\dots\}\rangle \not\sim^{\dagger} \langle\{\mathcal{M}'_2,\mathcal{M}'_4,\mathcal{M}'_6,\dots\},\{\mathcal{M}'_1,\mathcal{M}'_2,\mathcal{M}'_3,\dots\}\rangle$$

despite the existence of a countable uniform expansion. In contrast, the comparative confirmation ranking $\precsim^{\dagger}$ is fixed under n-uniform expansions of infinite sets.

Proposition 5.13

Let $\Omega_{\mathcal{L}^+}$ be an n-uniform model space expansion of $\Omega_{\mathcal{L}}$ for any $n \in \mathbb{N}^+$ and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then for any A, B from $\mathcal{F}^+_{\mathcal{L},\Sigma}$,

$$A \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B \text{ if and only if } A^+ \precsim^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$$

where A^+ and B^+ are the collections containing all and only those $\Omega_{\mathcal{L}^+,\Sigma}$ models which extend models in A and B respectively.

Unfortunately, extending this to cross-set comparisons still leads to undesirable results. Letting $\mathcal{M}_{R_k}$ be the canonical-domain model with a single element and non-empty unary relation R_k, we have

$$\langle\{\mathcal{M}_{R_0}\},\{\mathcal{M}_{R_0},\mathcal{M}_{R_1},\mathcal{M}_{R_2},\mathcal{M}_{R_3},\dots\}\rangle \sim \langle\{\mathcal{M}_{R_0}\},\{\mathcal{M}_{R_0},\mathcal{M}_{R_2},\mathcal{M}_{R_4},\mathcal{M}_{R_6},\dots\}\rangle$$

despite $\{\mathcal{M}_{R_0}\} \sim^{\dagger} \{\mathcal{M}_{R_0}\}$ and $\{\mathcal{M}_{R_0},\mathcal{M}_{R_2},\mathcal{M}_{R_4},\mathcal{M}_{R_6},\dots\} \prec^{\dagger} \{\mathcal{M}_{R_0},\mathcal{M}_{R_1},\mathcal{M}_{R_2},\mathcal{M}_{R_3},\dots\}$. While each individual uniform expansion appears proportion-preserving, transitivity quickly leads to equivalences that, by the lights of $\precsim^{\dagger}$, ought not be. Neither generalization of (U) is viable.

Instead, we stipulate the desired comparisons directly as a third and fourth axiom.

C. *Certain Validities*
$\langle\Omega_A,\Omega_A\rangle \sim \langle\Omega_B,\Omega_B\rangle$.

I. *Impossible Contradictions*
$\langle\emptyset,\Omega_A\rangle \sim \langle\emptyset,\Omega_B\rangle$.

Definition 5.6 Let Ω_A be a non-empty collection of canonical-domain models, and A a subcollection of Ω_A. Let $\precsim^{PUCI}$ denote the transitive closure of those comparisons required by (P) $\wedge$ (U) $\wedge$ (C) $\wedge$ (I) over pairs $\langle A, \Omega_A\rangle$.

The resulting cross-set confirmation relation $\precsim^{PUCI}$ is a strict extension of $\precsim^{PU}$ with a number of attractive features:

Proposition 5.14

Let Ω be the collection of all atomic possibilities, Ω_A and Ω_B non-empty subcollections, and A, B subcollections of Ω_A and Ω_B respectively. $\precsim^{PUCI}$ satisfies

C0. Nontriviality
$\langle \emptyset, \Omega\rangle \prec \langle \Omega, \Omega\rangle$.

C1a. Reflexivity
$\langle A, \Omega_A\rangle \precsim \langle A, \Omega_A\rangle$.

C2. Transitivity
If $\langle A, \Omega_A\rangle \precsim \langle B, \Omega_B\rangle$ and $\langle B, \Omega_B\rangle \precsim \langle C, \Omega_C\rangle$, then $\langle A, \Omega_A\rangle \precsim \langle C, \Omega_C\rangle$.

C3b. Boundedness
$\langle \emptyset, \Omega\rangle \sim \langle \emptyset, \Omega_A\rangle \precsim \langle A, \Omega_A\rangle \precsim \langle \Omega_A, \Omega_A\rangle \sim \langle \Omega, \Omega\rangle$.

$\mathbf{E}_{\precsim^\dagger}$. $\precsim^\dagger$ Extension
$\langle A, \Omega_A\rangle \precsim \langle A', \Omega_A\rangle$ *if and only if* $A \precsim^\dagger A'$.

$\mathbf{E}_{\prec^\dagger}$. $\prec^\dagger$ Extension
If either both $A \prec^\dagger_\Omega B$ and $\Omega_B \precsim^\dagger_\Omega \Omega_A$ or both $\emptyset \prec^\dagger_\Omega A \precsim^\dagger_\Omega B$ and $\Omega_B \prec^\dagger_\Omega \Omega_A$, then $\langle A, \Omega_A\rangle \prec \langle B, \Omega_B\rangle$.

FC. Finite Comparisons
If Ω_A and Ω_B are finite, then $\langle A, \Omega_A\rangle \precsim \langle B, \Omega_B\rangle$ if and only if $\mu_{\Omega_A}(A) \leq \mu_{\Omega_B}(B)$ where $\mu_{\Omega_A} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ and $\mu_{\Omega_B} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ are the unique representations of $\precsim^\dagger_{\Omega_A}$ and $\precsim^\dagger_{\Omega_B}$ in $\langle \mathbb{Q} \cap [0,1], +\rangle$.

ID. Infinitesimal Degrees
If A, B are nonempty and finite sets, Ω_A is infinite, and Ω_B is finite, then $\langle A, \Omega_A\rangle \prec \langle B, \Omega_B\rangle$.

Whether $\precsim^{PUCI}$ is a maximal solution is unclear. It does, however, suffice for the most plausible claims about degrees of confirmation, viz. those induced by representation of $\precsim^\dagger$ over finite sets in $\langle \mathbb{Q} \cap [0,1], +\rangle$ and simple generalizations thereof.

5.3 Conclusion

Over the course of the last century, subjective and probabilistic accounts of rational credence in the style of Ramsey (1931), de Finetti ([1937] 1980), and Savage (1972) have entirely supplanted accounts of confirmation. The arguments thought to secure these probabilistic accounts of rational credence, however, establish considerably less than advertised. The Dutch book argument, representation arguments, and gradational accuracy arguments all either outright assume rational credences are real-valued or assume rational credences satisfy properties characteristic of the real numbers, viz. comparability and the Archimedean property. As a result, these probabilistic accounts of rational credence establish, at best, that rational credences are real-valued only if they are probability functions.

At the same time, the objections thought to tell decisively against accounts of confirmation likewise fall short of the mark. Ramsey's skeptical response to confirmation is undercut by classical logic itself. Both propositional and first-order logic induce corresponding confirmation relations; skeptics need look no further to see their error. D'Alembert's riddle meanwhile rightly cast serious doubt on the ease with which classical accounts of probability purported to identify equally possible cases. The foundations of modern first-order logic suffice, however, to repair this shortcoming, delivering an account of confirmation for finite sets of atomic possibilities in the process.

While the nature and scope of the final objection to confirmation has been systematically understated, Bertrand's paradox suffices to rule out only probabilistic accounts of confirmation. Not only is the $\precsim^{\dagger}$-account of confirmation immune to Bertrand's paradox, but also it is the maximal resolution possible in an account of confirmation. Despite the comparative nature of the $\precsim^{\dagger}$-account of confirmation, meaningful talk of degrees of confirmation can also be recovered with a conservative extension to cross-set comparisons. The proposed account of confirmation thus survives all three of the objections which originally drove confirmation into disrepute without sacrificing either objectivity or normativity.

Finally, the $\precsim^{\dagger}$-account of confirmation dovetails with both the foundations of probability and the flaws noted in contemporary arguments for probabilism. Not only is the $\precsim^{\dagger}$-account of confirmation a generalization of the classical account of probability, but—under the proposed account—support of a claim φ by evidence Σ is probabilistic when the set of atomic models in question is finite. As a result, rational credences in these same circumstances are also probabilistic. In infinite contexts, however, support of a claim φ by evidence Σ satisfies neither comparability nor the Archimedean property and thus is not probabilistic. As a result, rational credences in infinite contexts are not probability functions. The $\precsim^{\dagger}$-account thus readily explains both the appeal and the shortcomings of probabilistic accounts of rational credence.

Bibliography

Baillie, Patricia. 1973. "Confirmation and the Dutch Book Argument." *The British Journal for the Philosophy of Science* 24:393–397.

Bangu, Sorin. 2009. "On Bertrand's paradox." *Analysis* 70:30–35.

Benci, Vieri, Horsten, Leon, and Wenmackers, Sylvia. 2013. "Non-archimedean probability." *Milan Journal of Mathematics* 81:121–151.

Bernoulli, Jakob. 1713. *Ars conjectandi*. Impensis Thurnisiorum, fratrum.

Bertrand, Joseph. 1889. *Calcul des probabilités*. Gauthier-Villars.

Bradley, Richard. 2017. *Decision theory with a human face*. Cambridge University Press.

Burgess, John P. 2003. "Which Modal Models are the Right Ones (for Logical Necessity)?" *Theoria: An International Journal for Theory, History and Foundations of Science* 145–158.

Caie, Michael. 2013. "Rational probabilistic incoherence." *Philosophical Review* 122:527–575.

Cantor, Georg. 1895. "Beiträge zur Begründung der transfiniten Mengenlehre." *Mathematische Annalen* 46:481–512.

Carnap, Rudolf. 1962. *Logical Foundations of Probability*. University of Chicago Press.

Christensen, David. 1996. "Dutch-book arguments depragmatized: Epistemic consistency for partial believers." *The Journal of Philosophy* 93:450–479.

Crupi, Vincenzo and Tentori, Katya. 2016. "Confirmation Theory." In *The Oxford Handbook of Probability and Philosophy*.

Daston, Lorraine. 1995. *Classical probability in the Enlightenment*. Princeton University Press.

de Finetti, Bruno. [1937] 1980. "Foresight: Its Logical Laws, its Subjective Sources (1937)." *Studies in Subjective Probability* 55–118.

De Finetti, Bruno. [1937] 1980. “Foresight: Its Logical Laws, its Subjective Sources (1937).” *Studies in Subjective Probability* 55–118.

de Finetti, Bruno. 1951. *La “logica del plausibile” secondo la concezione di polya*. Soc. Italiana per il Progresso delle Scienze.

—. 1979. *Theory of Probability. A Critical Introductory Treatment*.

Debreu, Gerard. 1954. “Representation of a preference ordering by a numerical function.” *Decision processes* 3:159–165.

Easwaran, Kenny and Fitelson, Branden. 2012. “An ‘Evidentialist’ Worry About Joyce’s Argument for Probabilism.” *Dialectica* 66:425–433.

Edwards, A. W. F. 1978. “Commentary on the Arguments of Thomas Bayes.” *Scandinavian Journal of Statistics* 5:116–118. ISSN 03036898, 14679469.

Elliott, Edward. 2019. “Betting against the Zen Monk: on preferences and partial belief.” *Synthese* ISSN 1573-0964. doi:10.1007/s11229-019-02308-4.

Ellis, Brian. 1960. “Some fundamental problems of direct measurement.” *Australasian Journal of Philosophy* 38:37–47.

Eriksson, Lina and Hájek, Alan. 2007. “What are Degrees of Belief?” *Studia Logica* 86:183–213.

Etchemendy, John. 1990. *The Concept of Logical Consequence*. Csli Publications.

Fine, Terrence L. 1973. *Theories of Probability: An Examination of Foundations*. Academic Press.

Friedman, Nir and Halpern, Joseph Y. 1995. “Plausibility Measures: A User’s Guide.” In *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence*, 175–184. Morgan Kaufmann Publishers Inc.

Galilei, Galileo. 1638. *Dialogues Concerning Two New Sciences*. Dover.

Gillies, Donald. 2000. *Philosophical theories of probability*. Routledge.

Greaves, Hilary. 2013. “Epistemic decision theory.” *Mind* 122:915–952.

Gyenis, Zalán and Rédei, Miklós. 2014. “Defusing Bertrand’s paradox.” *The British Journal for the Philosophy of Science* 66:349–373.

Hacking, Ian. 1971. “Equipossibility Theories of Probability.” *The British Journal for the Philosophy of Science* 22:339–355.

—. 2006. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge University Press.

Hájek, Alan. 2008. "Dutch Book Arguments." In Paul Anand, Prasanta Pattanaik, and Clemens Puppe (eds.), *The Oxford Handbook of Rational and Social Choice*. Oxford University Press.

Hájek, Alan. 2009. "Arguments For—Or Against—Probabilism?" In Franz Huber and Christoph Schmidt-Petri (eds.), *Degrees of Belief*, 229–251. Dordrecht: Springer Netherlands. ISBN 978-1-4020-9198-8. doi:10.1007/978-1-4020-9198-8_9.

—. 2010. "Staying Regular." *Unpublished manuscript* .

Hájek, Alan. 2012. "Interpretations of Probability." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition.

Hájek, Alan and Joyce, James M. 2008. "Confirmation." In S. Psillos and M. Curd (eds.), *The Routledge Companion to the Philosophy of Science*. Routledge.

Halpern, Joseph Y. 2001. "Plausibility measures: A general approach for representing uncertainty." In *International Joint Conference on Artificial Intelligence*, volume 17, 1474–1483.

—. 2017. *Reasoning About Uncertainty*. MIT press.

Hempel, Carl G. 1945. "Studies in the Logic of Confirmation (I.)." *Mind* 54:1–26.

Howson, Colin. 1991. "The 'Old Evidence' Problem." *The British Journal for the Philosophy of Science* 42:547–555.

Howson, Colin and Urbach, Peter. 2006. *Scientific reasoning: the Bayesian approach*. Open Court Publishing.

Jaynes, Edwin T. 1973. "The well-posed problem." *Foundations of Physics* 3:477–492.

Jeffrey, Richard. 1986. "Probabilism and induction." *Topoi* 5:51–58.

Jeffreys, Sir Harold. 1967. *Theory of Probability: 3d Ed.* Clarendon Press.

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65:575–603.

Joyce, James M. 2009. "Accuracy and coherence: Prospects for an alethic epistemology of partial belief." In *Degrees of belief*, 263–297. Springer.

Joyce, James M. 2010. "A Defense of Imprecise Credences in Inference and Decision Making 1." *Philosophical Perspectives* 24:281–323.

Kemeny, John G. 1955. "Fair bets and inductive probabilities." *The Journal of Symbolic Logic* 20:263–273.

Kendall, Maurice George and Morán, Patrick Alfred. 1963. *Geometrical probability*. Hafner Pub. Co.

Kennedy, Ralph and Chihara, Charles. 1979. "The Dutch book argument: Its logical flaws, its subjective sources." *Philosophical Studies* 36:19–33.

Keynes, John Maynard. 1921. *A Treatise on Probability*. Dover Publications.

Koopman, Bernard O. 1940. "The Bases of Probability." *Bulletin of the American Mathematical Society* 46:763–774.

Kraft, Charles H, Pratt, John W, and Seidenberg, Abraham. 1959. "Intuitive probability on finite sets." *The Annals of Mathematical Statistics* 30:408–419.

Krantz, David, Luce, Duncan, Suppes, Patrick, and Tversky, Amos. 1971. "Foundations of Measurement, Vol. I: Additive and Polynomial Representations." .

Lehman, R Sherman. 1955. "On confirmation and rational betting." *The Journal of Symbolic Logic* 20:251–262.

Levi, Isaac. 1985. "Imprecision and Indeterminacy in Probability Judgment." *Philosophy of Science* 52:390–409.

Lewis, David. 1981. "A Subjectivist's Guide to Objective Chance." In William L. Harper, Robert Stalnaker, and Glenn Pearce (eds.), *IFS: Conditionals, Belief, Decision, Chance and Time*, 267–297. Dordrecht: Springer Netherlands. ISBN 978-94-009-9117-0. doi: 10.1007/978-94-009-9117-0_14.

Lindley, Dennis V. 1982. "Scoring Rules and the Inevitability of." *International Statistical Review/Revue Internationale de Statistique* 50:1–11.

Luce, R Duncan. 1967. "Sufficient conditions for the existence of a finitely additive probability measure." *The Annals of Mathematical Statistics* 38:780–786.

Maher, Patrick. 1996. "Subjective and objective confirmation." *Philosophy of Science* 63:149–174.

—. 2002. "Joyce's argument for probabilism." *Philosophy of Science* 69:73–81.

Mancosu, Paolo. 2006. "Tarski on Models and Logical Consequence." In D. J. Ferreirós and J. Gray (eds.), *The architecture of modern mathematics*, 209–237. Oxford: Oxford University Press.

—. 2009. "Measuring the size of infinite collections of natural numbers: Was Cantor's theory of infinite number inevitable?" *The Review of Symbolic Logic* 2:612–646.

—. 2010. “Fixed- versus Variable-domain Interpretations of Tarski’s Account of Logical Consequence.” *Philosophy Compass* 5:745–759.

Marinoff, Louis. 1994. “A resolution of Bertrand’s paradox.” *Philosophy of Science* 61:1–24.

Mikkelson, Jeffrey M. 2004. “Dissolving the wine/water paradox.” *The British Journal for the Philosophy of Science* 55:137–145.

Milne, Peter. 1983. “A note on scale invariance.” *The British Journal for the Philosophy of Science* 34:49–55.

Myrvold, Wayne. 2019. *Beyond Chance and Credence*. Forthcoming.

Narens, Louis. 2007. *Theories of Probability: An Examination of Logical and Qualitative Foundations*, volume 2. World Scientific.

Norton, John D. 2008. “Ignorance and indifference.” *Philosophy of Science* 75:45–68.

Pettigrew, Richard. 2016. *Accuracy and the Laws of Credence*. Oxford University Press.

Pierre-Simon, Frederick Wilson, Laplace; *Trans. by* Truscott and Emory, Frederick Lincoln. [1814] 1951. *A Philosophical Essay on Probabilities*, volume 166. Dover Publications.

Poincaré, Henri. 1923. *Calcul des probabilitées*. Gauthier-Villiers.

Predd, Joel B, Seiringer, Robert, Lieb, Elliott H, Osherson, Daniel N, Poor, H Vincent, and Kulkarni, Sanjeev R. 2009. “Probabilistic coherence and proper scoring rules.” *IEEE Transactions on Information Theory* 55:4786–4792.

Ramsey, Frank P. 1931. “Truth and probability (1926).” *The Foundations of Mathematics and other Logical Essays* 156–198.

Reichenbach, Hans. 1971. *The theory of probability*. Univ of California Press.

Reid, Constance and Weyl, Hermann. 2013. *Hilbert*. Springer-Verlag.

Robert, Christian P, Chopin, Nicolas, Rousseau, Judith, et al. 2009. “Harold Jeffreys’s theory of probability revisited.” *Statistical Science* 24:141–172.

Rosenkrantz, Roger D. 1977. *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*. D. Reidel Publishing Company.

—. 1981. *Foundations and Applications of Inductive Probability*. Ridgeview Press.

Rowbottom, Darrell P. 2013. “Bertrand’s Paradox revisited: Why Bertrand’s ‘solutions’ are all inapplicable.” *Philosophia Mathematica* 21:110–114.

Rowbottom, Darrell P. and Shackel, Nicholas. 2010. “Bangu’s random thoughts on Bertrand’s paradox.” *Analysis* 70:689–692.

Savage, Leonard J. 1971. “Elicitation of personal probabilities and expectations.” *Journal of the American Statistical Association* 66:783–801.

—. 1972. *The Foundations of Statistics*. Courier Corporation.

Scott, Dana. 1964. “Measurement structures and linear inequalities.” *Journal of Mathematical Psychology* 1:233–247.

Shackel, Nicholas. 2007. “Bertrand’s Paradox and the Principle of Indifference.” *Philosophy of Science* 74:150–175.

Shimony, Abner. 1955. “Coherence and the Axioms of Confirmation.” *The Journal of Symbolic Logic* 20:1–28.

—. 1988. “An Adamite Derivation of the Principles of the Calculus of Probability.” In James H. Fetzer (ed.), *Probability and Causality: Essays in Honor of Wesley C. Salmon*, 79–89. Dordrecht: Springer Netherlands. ISBN 978-94-009-3997-4. doi: 10.1007/978-94-009-3997-4_3.

Skyrms, Brian. 1980. “Higher Order Degrees of Belief.” In D. H. Mellor (ed.), *Prospects for Pragmatism*, 109–137. Cambridge University Press.

Todhunter, Isaac. 1865. *A History of the Mathematical Theory of Probability: From the Time of Pascal to That of Laplace*. Macmillan and Company: London.

van Fraassen, Bas. 1983. “Calibration: A Frequency Justification for Personal Probability.” In R. Cohen and L. Laudan (eds.), *Physics, Philosophy, and Psychoanalysis*. D. Reidel.

—. 1989. *Laws and Symmetry*. Oxford University Press.

von Kries, J. 1886. *Die Principien der Wahrscheinlichkeitsrechnung*. Eine logische Untersuchung.

von Mises, Richard. 1957. *Probability, Statistics, and Truth*. Courier Corporation.

Weisberg, Jonathan. 2011. “Varieties of Bayesianism.” In Dov M. Gabbay, Stephan Hartmann, and John Woods (eds.), *Handbook of the History of Logic*.

Zabell, Sandy L. 1989. “The Rule of Succession.” *Erkenntnis* 31:283–321.

Appendices

Appendix A

Proofs for Chapter 2

A.1 Proofs for Section 2.1

Proposition A.1

Let O be a set and each of $\precsim, \prec, \sim$ binary relations over O. Then, the following are equivalent:

(1) (i) and (ii) hold;

(2) (iii) holds, $\sim$ is symmetric, and $\prec$ is exclusive.

Proof. ($\Rightarrow$)
($\Rightarrow$ for (iii))
Suppose that $o \precsim o'$. If $o' \precsim o$ as well, then $o \sim o'$ by (ii). If $o' \not\precsim o$, then $o \prec o'$ by (i). In either case, $o \prec o'$ or $o \sim o'$.
($\Leftarrow$ for (iii))
Suppose that $o \prec o'$ or $o \sim o'$. By (i) and (ii), this is equivalent to both $o \precsim o'$ and $o' \not\precsim o$ or both $o \precsim o'$ and $o' \precsim o$. This in turn is equivalent to $o \precsim o'$ as desired.

Symmetry and Exclusivity
Follow immediately from (i) and (ii).

($\Leftarrow$)
($\Rightarrow$ for (i))
Suppose $o \prec o'$. It follows from (iii) that $o \precsim o'$. Since $o \prec o'$ implies $o' \not\prec o$ and $o \nsim o'$, we likewise have that $o' \not\prec o$ and $o \nsim o'$. By (iii), this is equivalent to $o' \not\precsim o$.
($\Leftarrow$ for (i))
Suppose that $o \precsim o'$ and $o' \not\precsim o$. By (iii), either $o \prec o'$ or $o \sim o'$ and both $o' \not\prec o$ and $o' \nsim o$.

By $\sim$ a symmetric relation, $o \nsim o'$, and thus $o \prec o'$.

($\Rightarrow$ for (ii))
Suppose $o \sim o'$. It follows from (iii) that $o \precsim o'$. Since $\sim$ symmetric, we likewise have that $o' \sim o$, and thus $o' \precsim o$.
($\Leftarrow$ for (ii))
Suppose that $o \precsim o'$ and $o' \precsim o$. By (iii), either $o \prec o'$ or $o \sim o'$ and either $o' \prec o$ or $o' \sim o$. Since $o \prec o'$ implies $o' \nprec o$ and $o' \prec o$ implies $o \nprec o'$, it must be that $o \sim o'$ or $o' \sim o$. By $\sim$ symmetric, these latter are equivalent.

□

Proposition A.2

Let a set of objects O be given. Then,

(i) For any comparative ranking $\precsim$ over O, there exists an absolute scale $\langle A, \leq \rangle$ and absolute ranking $\mu : O \to A$ such that

$$o \precsim o' \Leftrightarrow \mu(o) \leq \mu(o').$$

(ii) For any absolute scale $\langle A, \leq \rangle$ and absolute ranking $\mu : O \to A$, there exists a comparative ranking $\precsim$ over O such that

$$o \precsim o' \Leftrightarrow \mu(o) \leq \mu(o').$$

Proof. (i)
Let an object set O and comparative ranking $\precsim$ over O be given. Let A be the equivalence classes of O under $\sim$ and let $\mu : O \to A$ be defined by $\mu(o) = [o]_\sim$. Finally, define $\leq$ over A by:

$$[o]_\sim \leq [o']_\sim \Leftrightarrow o \precsim o'.$$

(ii)
Define $\precsim$ over O by $o \precsim o' \Leftrightarrow \mu(o) \leq \mu(o')$. Reflexivity and transitivity for $\precsim$ follow immediately from the reflexivity and transitivity conditions on the absolute scale. □

A.2 Proofs for Section 2.2

Proposition A.3

Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a likelihood assignment $\mu : L \rightarrow \mathbb{P}$ satisfying (N), (C), and (I), μ satisfies (F) only if $\circ$ satisfies commutativity, additive identity, and existence of complements over $\mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}$ where $\circ : \mathbb{P} \times \mathbb{P} \twoheadrightarrow \mathbb{P}$ is defined by

$$\mu(\varphi) \circ \mu(\psi) = \mu((\varphi \vee \psi))$$

for $\varphi, \psi \in L$ inconsistent.

Proof. Suppose that L is a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ is a model space for L, and μ is a likelihood assignment $\mu : L \rightarrow \mathbb{P}$ satisfying *Normativity of Logic*, *Certain Validities*, and *Impossible Contradictions*. Define

$$\mu(\varphi) \circ \mu(\psi) = \mu((\varphi \vee \psi))$$

for $\varphi, \psi \in L$ inconsistent. By *Functionality for Disjoint Disjunction*, $\circ$ is a well-defined partial function from $\mathbb{P} \times \mathbb{P}$ to $\mathbb{P}$.

Since $(\varphi \vee \psi)$ is logically equivalent to $(\psi \vee \varphi)$, *Normativity of Logic* requires that $\mu((\varphi \vee \psi)) = \mu((\psi \vee \varphi))$ for $\varphi, \psi \in L$ inconsistent. By construction, we thus have that

$$\mu(\varphi) \circ \mu(\psi) = \mu(\psi) \circ \mu(\varphi).$$

$\circ$ thus satisfies *Commutativity*.

Since $(\varphi \vee \bot)$ is logically equivalent to φ, *Normativity of Logic* requires that $\mu((\varphi \vee \bot)) = \mu(\varphi)$ for any $\varphi \in L$. By the construction of $\circ$ together with *Impossible Contradictions*, we thus have that

$$\mu(\varphi) \circ \dot{0} = \mu(\varphi)$$

for all $\varphi \in L$. $\circ$ thus satisfies *Additive Identity* over $\mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}$.

Since $(\varphi \vee \neg phi)$ is logically equivalent to $\top$, *Normativity of Logic* together with *Certain Validities* requires that $\mu((\varphi \vee \neg\varphi)) = \dot{1}$ for any $\varphi \in L$. Since φ and $\neg\varphi$ are inconsistent, we thus have that

$$\mu(\varphi) \circ \mu(\neg\varphi) = \dot{1}$$

for all $\varphi \in L$. $\circ$ then satisfies *Existence of Complements* over $\mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}$.

∎

Proposition A.4

Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$ satisfying (N), (C), (I), and (M), μ also satisfies

D. Decomposition
For any disjoint $A_1, B_1 \in \mathcal{F}_{\mathcal{L}}$ and disjoint $A_2, B_2 \in \mathcal{F}_{\mathcal{L}}$ such that $\mu(A_1) \leq \mu(A_2)$ and $\mu(B_1) \leq \mu(B_2)$, $\mu(A_1 \cup B_1) \leq \mu(A_2 \cup B_2)$.

F. Functionality for Inconsistent Disjunction / Disjoint Union
For any two pairs of inconsistent sentences φ, ψ and φ', ψ', if $\mu(\varphi) = \mu(\varphi')$ and $\mu(\psi) = \mu(\psi')$, then $\mu((\varphi \vee \psi)) = \mu((\varphi' \vee \psi'))$.

Proof. We prove (D) first. Let $A_1, B_1 \in \mathcal{F}_{\mathcal{L}}$ disjoint and $A_2, B_2 \in \mathcal{F}_{\mathcal{L}}$ disjoint such that $\mu(A_1) \leq \mu(A_2)$, $\mu(B_1) \leq \mu(B_2)$. Using (M) on the first inequality, $\mu(A_1 \cup (B_1 - A_2)) \leq \mu(A_2 \cup (B_1 - A_2))$, and thus $\mu(A_1 \cup (B_1 - A_2)) \leq \mu(A_2 \cup B_1)$. Using (M) on the second inequality, $\mu(B_1 \cup (A_2 - B_1)) \leq \mu(B_2 \cup (A_2 - B_1))$, and thus $\mu(A_2 \cup B_1) \leq \mu(B_2 \cup (A_2 - B_1))$. By transitivity, $\mu(A_1 \cup (B_1 - A_2)) \leq \mu(B_2 \cup (A_2 - B_1))$. Finally, applying (M) one last time gives $\mu(A_1 \cup (B_1 - A_2) \cup (B_1 \cap A_2)) \leq \mu(B_2 \cup (A_2 - B_1) \cup (B_1 \cap A_2))$, and thus $\mu(A_1 \cup B_1) \leq \mu(A_2 \cup B_2)$.

(F) now follows by two applications of (D). □

Proposition A.5

For a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$,

(i) (N), (C), (I) $\not\Rightarrow$ (M).

(ii) (M), (C), (I) $\not\Rightarrow$ (N).

(iii) (N), (M), (I) $\not\Rightarrow$ (C).

(iv) (N), (C), (M) $\not\Rightarrow$ (I).

Proof.
(i) follows from the example showing that (N), (C), (I), (F), (A) $\not\Rightarrow$ (M).

For (ii), consider the propositional language generated from $\sigma = \{P, Q\}$ with $\Omega_{\mathcal{L}} = \{v_1, v_2, v_3, v_4\}$ the set of all propositional models for L. Set $\mu(\emptyset) = 0$, $\mu(\Omega_{\mathcal{L}}) = 1$, and all remaining sets from $\mathcal{F}_{\mathcal{L}}$ incomparable between the two. It's trivial to verify that (C), (I), (M), and ¬(N).

For (iii), consider the propositional language generated from $\sigma = \{P\}$ with $\Omega_{\mathcal{L}} = \{v_1, v_2\}$ the set of all propositional models for L. Set $\mu(\emptyset) = 0$, $\mu(\{v_1\}) = \frac{1}{3}$, $\mu(\{v_2\}) = \frac{1}{3}$, and $\mu(\{v_1, v_2\}) = \frac{2}{3}$ but $\dot{1}$ to 1. It's trivial to verify that (N), (I), (M), and ¬(C).

For (iv), consider the propositional language generated from $\sigma = \{P\}$ with $\Omega_{\mathcal{L}} = \{v_1, v_2\}$ the set of all propositional models for L. Set $\mu(\emptyset) = \frac{1}{3}$, $\mu(\{v_1\}) = \frac{2}{3}$, $\mu(\{v_2\}) = \frac{2}{3}$, and $\mu(\{v_1, v_2\}) = 1$ but $\dot{0}$ to 0. It's trivial to verify that (N), (C), (M), and ¬(I).

□

Proposition A.6

Given a formal language L with Boolean connectives, a model space $\Omega_{\mathcal{L}}$ for L, and a likelihood assignment $\mu : L \to \mathbb{P}$ satisfying (N), (C), (I), (A), and (M), $\langle \mathbb{P}|_{\mu[\mathcal{F}_{\mathcal{L}}]}, \circ \rangle$ is a likelihood structure where $\circ : \mathbb{P} \times \mathbb{P} \twoheadrightarrow \mathbb{P}$ is defined by

$$\mu(\varphi) \circ \mu(\psi) = \mu((\varphi \vee \psi))$$

for $\varphi, \psi \in L$ inconsistent.

Proof. Conditions (a)-(c) follow from (F) and thus (M). Condition (d) on $\circ$ follows immediately from (A); condition (e) on $\circ$ follows immediately from (M). □

A.3 Proofs for Section 2.3

Proposition A.7

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, $\mathbb{P}$ a likelihood space, $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ an NCI-likelihood assignment. Then, the relation $\precsim_{\mu}$ over $\mathcal{F}_{\mathcal{L}}$ defined by

$$A \precsim_{\mu} B \Leftrightarrow \mu(A) \leq \mu(B)$$

satisfies

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\perp \precsim A \precsim \Omega_{\mathcal{L}}$.

CN. Normativity of Logic
If $A \subseteq B$, then $A \precsim B$.

Proof. (C0) follows from restriction (iv.a) on likelihood spaces together with both restriction (I) and restriction (C) on likelihood assignments. (C1a) and (C2) follow immediately given the definition of the $\precsim$-relation and restrictions (iv.b) and (iv.c) on likelihood spaces. The first part of (C3b) is guaranteed by restriction (I) on likelihood assignments together with restriction (iv.e) on likelihood spaces; the second part of (C3b) is then delivered by restriction (C) on likelihood assignments and restriction (iv.e) on likelihood spaces. Finally, (CN) is guaranteed by restriction (N) on likelihood assignments. $\square$

Proposition A.8

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

CN. Normativity of Logic
If $A \subseteq B$, then $A \precsim B$.

Then, there exists a likelihood space $\mathbb{P}_{\precsim}$ ("the likelihood space induced by $\precsim$") such that $\precsim$ is representable in $\mathbb{P}_{\precsim}$ by an NCI-Likelihood Assignment.

Proof. Let $\mathcal{F}_{\mathcal{L}}$ and $\precsim$ be given. Take the domain of $\mathbb{P}_{\precsim}$ to be $\mathcal{F}_{\mathcal{L}}/\sim$ (where $A \sim B$ iff $A \precsim B$ and $B \precsim A$) with $\dot{0}$ as $[\emptyset]$, $\dot{1}$ as $[\Omega_{\mathcal{L}}]$, and $\leq$ the partial ordering induced by $\precsim$. It is easy to verify that $\mathbb{P}_{\precsim}$ is a likelihood space. Let $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}_{\precsim}$ merely map any $A \in \mathcal{F}_{\mathcal{L}}$ to $[A] \in \mathbb{P}_{\precsim}$. μ clearly satisfies (C) and (I). Suppose that $\varphi \models \psi$. Then, $[\![\varphi]\!] \subseteq [\![\psi]\!]$. By (CN), $[\![\varphi]\!] \precsim [\![\psi]\!]$. It follows immediately by the construction of μ and $\mathbb{P}_{\precsim}$ that $\mu([\![\varphi]\!]) \leq \mu([\![\psi]\!])$. μ therefore satisfies (N) as well. □

Proposition A.9

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$. If $\precsim$ is representable in a likelihood space $\mathbb{P}$ and $\mathbb{P}$ embeds into another likelihood space $\mathbb{P}^+$, then $\precsim$ is representable in $\mathbb{P}^+$. Further, if $\precsim$ is representable in a likelihood structure $\langle \mathbb{P}, \circ \rangle$ and $\langle \mathbb{P}, \circ \rangle$ embeds into another likelihood structure $\langle \mathbb{P}^+, \circ^+ \rangle$, then $\precsim$ is representable in $\langle \mathbb{P}^+, \circ^+ \rangle$.

Proof. Simply rewrite the representing likelihood assignment $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ into a function $\mu^+ : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}^+$ with no change in the values assigned. It is trivial to verify that μ^+ is a likelihood assignment which represents $\precsim$ in the extended space. □

Proposition A.10

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

CN. Normativity of Logic
If $A \subseteq B$, then $A \precsim B$.

Then, $\precsim$ is representable in a likelihood space $\mathbb{P}$ by an NCI-likelihood assignment if and only if $\mathbb{P}_{\precsim}$ embeds into $\mathbb{P}$.

Proof.
($\Rightarrow$)
Suppose that $\precsim$ is representable in $\mathbb{P}$ by an NCI-likelihood assignment. By definition, there exists $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ such that for any $A, B \in \mathcal{F}_{\mathcal{L}}$, $A \precsim B$ if and only if $\mu(A) \leq \mu(B)$. Define $\delta : \mathbb{P}_{\precsim} \to \mathbb{P}$ by $\delta([A]) = \mu(A)$. Note that δ is well-defined since, by above, any two representatives of the same equivalence class under $\sim$ must be assigned the same likelihood by μ. It's easy to verify both that $\delta'(\dot{0}^{\mathbb{P}_{\precsim}}) = \dot{0}^{\mathbb{P}}$ and $\delta(\dot{1}^{\mathbb{P}_{\precsim}}) = \dot{1}^{\mathbb{P}}$ using (I) and (C). Finally, note that $\langle [A], [B] \rangle \in \leq^{\mathbb{P}_{\precsim}}$ if and only if $\langle \delta([A]), \delta([B]) \rangle \in \leq^{\mathbb{P}}$ by the construction of δ and μ representing.

($\Leftarrow$)
Suppose that $\mathbb{P}_{\precsim}$ embeds into $\mathbb{P}$. Let $\delta : \mathbb{P}_{\precsim} \to \mathbb{P}$ be this embedding. Define $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ by $\mu(A) = \delta([A])$. By δ an embedding and $\mu' : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}_{\precsim}$ defined by $\mu'(A) = [A]$ an NCI-likelihood assignment, it's easy to verify both that μ is an NCI-likelihood assignment and that μ represents $\precsim$ in $\mathbb{P}$. $\square$

Proposition A.11

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, $\mathbb{P}$ a likelihood space, $\mu : \mathcal{F}_{\mathcal{L}} \to \mathbb{P}$ an NCIAM-likelihood assignment. Then, the relation $\precsim_\mu$ over $\mathcal{F}_{\mathcal{L}}$ defined by

$$A \precsim_\mu B \Leftrightarrow \mu(A) \leq \mu(B)$$

satisfies

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\perp \precsim A \precsim \Omega_{\mathcal{L}}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, and $G_4 \sim F_3 \cup E_3$ with matching subscripts disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Proof. It remains only to establish (C4) and (CA). Note that (C4) is entailed by (M) given the definition of $\precsim_\mu$ while (CA) is entailed by (A). □

Proposition A.12

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_{\mathcal{L}}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_{\mathcal{L}}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, $G_4 \sim F_3 \cup E_3$, and all sets with matching subscripts are disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Then, there exists a likelihood structure $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ ("the likelihood structure induced by $\precsim$") such that $\precsim$ is representable in $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ by an NCIAM-Likelihood Assignment.

Proof. Let $\mathcal{F}_{\mathcal{L}}$ and $\precsim$ be given. Define $\mathbb{P}_{\precsim}$ as usual and $\circ_{\precsim} : \mathbb{P}_{\precsim} \times \mathbb{P}_{\precsim} \twoheadrightarrow \mathbb{P}_{\precsim}$ by

$$[A] \circ_{\precsim} [B] \downarrow= [C] \text{ if and only if there exists disjoint } A', B' \in \mathcal{F}_{\mathcal{L}} \text{ such that } A' \in [A], \\ B' \in [B], \text{ and } A' \cup B' \in [C].$$

Noting that (C3b) and (C4) entail (CN), $\mathbb{P}_{\precsim}$ is a likelihood space by the previous subsection. We know verify that $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ is a likelihood structure.

(a) Suppose that $[A] \circ_{\precsim} [B] \downarrow= [C]$. By definition, there exists disjoint sets $A', B' \in \mathcal{F}_{\mathcal{L}}$ such that $A' \in [A], B' \in [B]$, and $A' \cup B' \in [C]$. It follows immediately that there exists disjoint sets $A', B' \in \mathcal{F}_{\mathcal{L}}$ such that $B' \in [B], A' \in [A]$, and $B' \cup A' \in [C]$, and thus that $[B] \circ_{\precsim} [A] \downarrow= [A] \circ_{\precsim} [B]$.

(b) Suppose that $[A] \in \mathbb{P}_{\precsim}$. Then, $A, \emptyset$ are disjoint sets such that $A \cup \emptyset = A$. It follows that $[A] \circ_{\precsim} \dot{0} \downarrow= [A]$.

(c) Suppose that $[A] \in \mathbb{P}_{\precsim}$. Then, A, A^c are disjoint sets such that $A \cup A^c = \Omega_{\mathcal{L}}$. It follows that $[A] \circ_{\precsim} [A^c] \downarrow= \dot{1}$.

(d) Follows directly from (CA).

(e) Follows directly from (C4).

Let $\mu : \mathcal{F}_\mathcal{L} \to \mathbb{P}_\precsim$ merely map any $A \in \mathcal{F}_\mathcal{L}$ to $[A] \in \mathbb{P}_\precsim$. μ clearly satisfies (C) and (I). Suppose that $\varphi \models \psi$. Then, $[\![\varphi]\!] \subseteq [\![\psi]\!]$. By (C3a), $\emptyset \precsim [\![\psi]\!] - [\![\varphi]\!]$ and thus by (C4) $[\![\varphi]\!] \precsim [\![\psi]\!]$. It follows immediately by the construction of μ and $\mathbb{P}_\precsim$ that $\mu([\![\varphi]\!]) \leq \mu([\![\psi]\!])$. μ therefore satisfies (N) as well. Finally, μ applies $\circ_\precsim$ for disjoint unions by construction. □

Proposition A.13

Let L be a formal language with Boolean connectives, $\Omega_\mathcal{L}$ a model space for L, and $\precsim$ a binary relation on $\mathcal{F}_\mathcal{L}$ satisfying

C0. Nontriviality
$\emptyset \prec \Omega_\mathcal{L}$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3b. Boundedness
$\bot \precsim A \precsim \Omega_\mathcal{L}$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, $G_4 \sim F_3 \cup E_3$, and all sets with matching subscripts are disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Then, $\precsim$ is representable in $\langle \mathbb{P}', \circ' \rangle$ if and only if $\langle \mathbb{P}_\precsim, \circ_\precsim \rangle$ embeds into $\langle \mathbb{P}', \circ' \rangle$.

Proof.
($\Rightarrow$)
Suppose that $\precsim$ is representable in $\langle \mathbb{P}', \circ' \rangle$. By definition, there exists $\mu : \mathcal{F}_\mathcal{L} \to \mathbb{P}$ such that for any $A, B \in \mathcal{F}_\mathcal{L}$, $A \precsim B$ if and only if $\mu(A) \leq \mu(B)$ and for any disjoint $C, D \in \mathcal{F}_\mathcal{L}$, $\mu(C \cup D) = \mu(C) \circ' \mu(D)$. Define $\delta : \mathbb{P}_\precsim \to \mathbb{P}$ by $\delta([A]) = \mu(A)$. Note that δ is well-defined since, by above, any two representatives of the same equivalence class under $\sim$ must be assigned the same likelihood by μ. It's easy to verify both that $\delta'(\dot{0}^{\mathbb{P}_\precsim}) = \dot{0}^{\mathbb{P}}$ and $\delta(\dot{1}^{\mathbb{P}_\precsim}) = \dot{1}^{\mathbb{P}}$ using (I) and (C). Note next that $\langle [A], [B] \rangle \in \leq^{\mathbb{P}_\precsim}$ if and only if $\langle \delta([A]), \delta([B]) \rangle \in \leq^{\mathbb{P}}$ by the construction of δ and μ representing. Finally, let $C, D \in \mathcal{F}_\mathcal{L}$ and disjoint. Then,

$\delta([C] \circ_{\precsim} [D]) = \delta([C \cup D]) = \mu(C \cup D) = \mu(C) \circ' \mu(D)$.

($\Leftarrow$)
Suppose that $\langle \mathbb{P}_{\precsim}, \circ_{\precsim} \rangle$ embeds into $\langle \mathbb{P}', \circ' \rangle$. Let $\delta : \mathbb{P}_{\precsim} \rightarrow \mathbb{P}'$ be this embedding. Define $\mu : \mathcal{F}_{\mathcal{L}} \rightarrow \mathbb{P}'$ by $\mu(A) = \delta([A])$. By δ an embedding and $\mu' : \mathcal{F}_{\mathcal{L}} \rightarrow \mathbb{P}_{\precsim}$ defined by $\mu'(A) = [A]$ representing $\precsim$, it's easy to verify that μ represents $\precsim$ in $\langle \mathbb{P}', \circ' \rangle$. $\square$

Appendix B

Proofs for Chapter 3

B.1 Proofs for Section 3.2

Proposition B.1

Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\precsim$ a binary relation over $\mathcal{F}$. Then,

(i) (C1b) $\Rightarrow$ (C1a)

(ii) (C5$_S$) $\Rightarrow$ (C2)

(iii) (C3a), (C5$_S$) $\Rightarrow$ (C3b)

(iv) (C5$_S$) $\Rightarrow$ (C4).

Proof.

(i) Trivial.

(ii) Suppose $A \precsim B$ and $B \precsim C$. Then, $\langle A, B, C\rangle$ and $\langle B, C, A\rangle$ are sequences which contain each possibility in Ω the same number of times and have both $A \precsim B$ and $B \precsim C$. By (C5$_S$), we thus have that $A \precsim C$ as desired.

(iii) Let $A \in \mathcal{F}$. $\langle \emptyset, \Omega\rangle$ and $\langle A^c, A\rangle$ are sequences which contain each possibility in Ω the same number of times. Further, $\emptyset \precsim A^c$ by C3a. By (C5$_S$), we thus have that $A \precsim \Omega$, and thus by another application of (C3a), $\emptyset \precsim A \precsim \Omega$.

(iv) Suppose that $A \cap C = B \cap C = \emptyset$.

($\Rightarrow$) Suppose $A \precsim B$. Then, $\langle A, B \cup C\rangle$ and $\langle B, A \cup C\rangle$ are sequences which contain each possibility in Ω the same number of times with $A \precsim B$. By (C5$_S$), $A \cup C \precsim B \cup C$.

($\Leftarrow$) Suppose $A \cup C \precsim B \cup C$. Then, $\langle A \cup C, B\rangle$ and $\langle B \cup C, A\rangle$ are sequences which contain each possibility in Ω the same number of times with $A \cup C \precsim B \cup C$. By (C5$_S$), $A \precsim B$.

□

Proposition B.2

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\precsim$ a binary relation on $\mathcal{F}_{\mathcal{L}}$. $\precsim$ is representable by an NCI-likelihood assignment into $\langle \mathbb{Q} \cap [0,1], 0, 1, \leq, +\rangle$ if and only if $\precsim$ satisfies C01b3a5$_S$.

Proof. The forward direction follows merely in virtue of noticing that the functions under consideration are a strict subset of those in the Scott (1964) theorem. For the converse, note that μ is completely determined once each individual model in $\Omega_{\mathcal{L}}$ has been assigned a value ($\Omega_{\mathcal{L}}$ is finite and the eventual likelihood assignment μ must be additive). Our strategy is simply to nudge any models assigned an irrational value onto a nearby rational one while maintaining normalization.

Let μ' be the real-valued function guaranteed by Scott's result. We now construct the desired function μ by modifying the probability of any model which is assigned an irrational value under μ'. Noting that there are only a finite number of non-equivalent sentences in L under $\Omega_{\mathcal{L}}$, there must be a minimum, non-zero value r such that $r = |\mu'(\varphi) - \mu'(\varphi')|$ for $\varphi, \varphi' \in \mathcal{L}$. Let q be a rational value such that $0 < q < r$. Next, let $I \subseteq \Omega_{\mathcal{L}}$ be the set of all models assigned irrational values by μ'. If $|I| = 0$, we may simply take $\mu = \mu'$ and be done. Suppose, then, that $|I| > 0$. For all $w \in \Omega_{\mathcal{L}} - I$, we set $\mu(w) = \mu'(w)$.

Treating models in I assigned the same probability by μ' as a group, we proceed through the groups (there must be at least two given that the entries of I are irrational and the sum rational) in order of decreasing probability relative to μ'. For all groups in I besides the last, set all w in a given group to the same rational value from the interval

$$(\mu'(w), \mu'(w) + \left(\frac{1}{2^{|\Omega_{\mathcal{L}}|} + 1}\right)q].$$

For every w in the last group G, set

$$\mu(w) = \frac{1 - \sum_{w' \in \Omega_{\mathcal{L}} - G} \mu(w')}{|G|}.$$

It is clear from the construction that $\mu(\emptyset) = 0$, that $\mu(\Omega_{\mathcal{L}}) = 1$, that μ is additive, and that μ assigns only rational values; μ is thus an NCI-likelihood assignment to $\langle \mathbb{Q} \cap [0,1], 0, 1, \leq, + \rangle$. Similarly, it is obvious that μ has maintained equality relationships within $\Omega_{\mathcal{L}} - I$ and within each group in I relative to that group. Notice, moreover, that none of the groups in I whose probability was increased could now be equal either to another such group or to a model in $\Omega_{\mathcal{L}} - I$ (q wouldn't be less than the smallest difference between any two sets from $\mathcal{F}$ relative to μ'). Finally, the last group in I can be equal neither to any of the previously adjusted groups (it had the lowest probability and was revised downward), nor any model in $\Omega_{\mathcal{L}} - I$ (q again wouldn't be less than the smallest difference between any two sets from $\mathcal{F}$ relative to μ'). It follows that μ maintains all equalities over $\Omega_{\mathcal{L}}$.

Lastly, notice that for any model w,

$$|\mu(w) - \mu'(w)| < \left(\frac{|\Omega_{\mathcal{L}}| - 1}{2^{|\Omega_{\mathcal{L}}|} + 1}\right) q$$

Since any $A \in \mathcal{F}_{\mathcal{L}}$ contains no more than $|\Omega_{\mathcal{L}}|$ models,

$$|\mu(A) - \mu'(A)| < \left(\frac{2^{|\Omega_{\mathcal{L}}|}}{2^{|\Omega_{\mathcal{L}}|} + 1}\right) q$$

And so, the largest possible change going from μ' to μ for any $A, B \in \mathcal{F}_{\mathcal{L}}$ relative to one another is strictly less than q which, by definition, is less than the smallest non-zero difference between any two sets under μ'. It follows immediately that μ preserves not only all equality relations over $\mathcal{F}_{\mathcal{L}}$, but also all $\prec$ relations as well. $\square$

Proposition B.3

Let L be a formal language, $\Omega_{\mathcal{L}}$ a model space for L which contains only a single model, and $\precsim$ a binary relation over $\mathcal{F}_{\mathcal{L}}$. Then, $\precsim$ satisfies C01b3a5$_S$ if and only if it is representable in

$$\langle \{0,1\}, 0, 1, \leq \rangle \textit{ with } \circ(x,y) = \begin{cases} \uparrow & \textit{if } x = 1 \textit{ and } y = 1 \\ \max(x,y) & \textit{otherwise.} \end{cases}$$

Proof.
($\Rightarrow$)
$\mathcal{F}_{\mathcal{L}}$ is always $\{\emptyset, \Omega_{\mathcal{L}}\}$. Set $\mu(\emptyset) = \dot{0}$ and $\mu(\Omega_{\mathcal{L}}) = \dot{1}$.

($\Leftarrow$)
$\mathcal{F}_{\mathcal{L}}$ is always $\{\emptyset, \Omega_{\mathcal{L}}\}$. We must have $\mu(\emptyset) = \dot{0}$ and $\mu(\Omega_{\mathcal{L}}) = \dot{1}$. It follows by definition that

$\emptyset \prec \Omega_{\mathcal{L}}$. C01b3a5$_S$ are all easy to verify. □

Proposition B.4

Let L be a formal language with Boolean connectives and $\Omega_{\mathcal{L}}$ a model space for L. Then, there exists a binary relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ satisfying C01b3a5$_S$ which is not representable in $\langle [0,1], 0, 1, \leq, + \rangle$.

Proof. It is easy to show that there exist linear orders whose cardinality is strictly greater than the continuum. Let $\langle A, < \rangle$ be such an ordering and set $\sigma = \{P_a \ : \ a \in A\}$ with the model space which contains only valuations which assign a single 'T' value. Given any sentence $\varphi \in \mathcal{L}$ can be expressed as a minimal disjunction of conjunctions of literals and every set in $\mathcal{F}_{\mathcal{L}}$ corresponds to some such sentence, we may define our preorder over $\mathcal{F}_{\mathcal{L}}$ by leveraging the syntactic structure of these disjunctions of conjunctions of literals.

In general, our strategy is simply to take a lexicographic ordering based on $\langle A, < \rangle$ of the sets in $\mathcal{F}_{\mathcal{L}}$ (note that every set either contains a finite number of valuations or is missing a finite number of valuations). Define $\precsim$ over $\mathcal{F}_{\mathcal{L}}$ by

- For any $A \in \mathcal{F}_{\mathcal{L}}$, $A \precsim A$
- For any P_a, $\emptyset \prec [\![P_a]\!]$
- $[\![P_a]\!] \prec [\![P_{a'}]\!]$ if and only if $a < a'$
- For any P_a and $P_{a'}$, $[\![P_a]\!] \prec [\![\neg P_{a'}]\!]$
- For any P_a and $P_{a'}$, $[\![\neg P_a]\!] \prec [\![\neg P_{a'}]\!]$ if and only if $a' < a$
- For any $\neg P_a$, $[\![\neg P_a]\!] \prec \Omega_{\mathcal{L}}$

That is, each propositional letter identifies a unique valuation and is ordered according to the associated $a \in A$. Moreover, all negative literals are ordered above all positive ones with the former reversing the order of the latter; this gives a total ordering over both the singletons of $\mathcal{F}_{\mathcal{L}}$ and every set missing only a single valuation.

Note next that intersections between these sets are only nontrivial for negative literals, i.e. sets missing only a single valuation. Given a finite intersection of such sets A, a finite number of valuations are missing. Ordering these valuations from greatest to least, set $A \prec \Omega_{\mathcal{L}} - \{v_1\}$ where v_1 is the greatest of these valuations (note the inversion caused because we are subtracting valuations). Further, for any valuation v^+ greater than v_1, set $\Omega_{\mathcal{L}} - \{v^+\} \prec A$. Finally, for any two finite intersections of such sets A and B, set $A \prec B$ if and only if, ordering the missing valuations from greatest to least, there exists a position in which the nth missing valuation of A is strictly less than the nth missing valuation of B

and all previous positions are equal, treating the lack of any additional missing valuations as greater than any valuation. It follows immediately that all conjunctions of literals (and any sentences equivalent to one of these) in $\mathcal{L}$ have been assigned a position in our ordering.

We need now only concern ourselves with disjunctions/unions of the sets above and, again, we turn to a lexicographic ordering. For any union A, order the union sets from the greatest A_1 to least A_n. Thus, $A = \cup_{i=1}^{n} A_i$, and we set both $A_1 \prec A$ and $A \prec A_1^+$ where A_1^+ is any non-unioned set greater than A_1 from the previous stage. For any two unioned sets A and B, set $A \prec B$ if and only if ordering each of the unions from greatest to least and starting with the former, there exists an n such that the nth greatest set of B is strictly greater than the nth greatest set of A with all previous disjuncts equal, treating the lack of any remaining sets as less than any actual set.

It is easy to verify that C0, C1a, Clb, C2, and C3 hold with the defined relation $\precsim$. Suppose that two sequences from $\mathcal{F}_{\mathcal{L}}$, $\langle A_1, \ldots, A_N \rangle$ and $\langle B_1, \ldots, B_N \rangle$, contain each point in $\Omega_{\mathcal{L}}$ the same number of times. Note that, without loss of generality, every A_i and B_i can be expressed as a union of intersections of our basic sets (singletons or the complements thereof). Assume that $A_i \precsim B_i$ for all $i < N$ and—for contradiction—that $A_N \prec B_N$. Let D denote the largest unioned set occurring within any of $B_1, \ldots, B_N$ which doesn't immediately reoccur in the corresponding A_i. Note that, if no such D exists, the two sequences contain equivalent sentences at each position, and thus $B_N \precsim A_N$ contra our supposition. Since the two sequences $\langle A_1, \ldots, A_N \rangle$ and $\langle B_1, \ldots, B_N \rangle$ contain the same number of truths, it must be that D also occurs within one of the $A_1, \ldots, A_N$ without appearing in the corresponding B_i. Consider this A_i and B_i. By the definition of D, there is no greater disjunct within B_i which doesn't also occur within A_i. But, the construction of $\precsim$ together with D's appearance in A_i and not in B_i entails that $B_i \prec A_i$—a contradiction. It follows immediately that C5a likewise holds for the constructed relation $\precsim$, and thus also C4.

Lastly, we show that no function μ represents $\precsim$ in $\langle [0,1], 0, 1, \leq, + \rangle$. Note that, by the construction of $\mathcal{L}$ and the mimicking of the $\langle A, \leq \rangle$ order within the $[\![P_a]\!]$, $\precsim$ makes each $[\![P_a]\!]$ for $a \in A$ distinct from every other. Since $|A| > \mathfrak{c}$, there exists, for any $\mu : \mathcal{L} \to [0,1]$, $a, a' \in A$ such that $a < a'$ (and thus $[\![P_a]\!] \prec [\![P_{a'}]\!]$), but $\mu(P_a) = \mu(P_{a'})$. □

Proposition B.5

Let $\langle \mathbb{P}, \circ \rangle$ be a likelihood structure. Then, there exists a propositional language L, model space $\Omega_{\mathcal{L}}$, and C01b3a5$_S$ relation $\precsim$ on $\mathcal{F}_{\mathcal{L}}$ such that $\precsim$ is not represented in $\langle \mathbb{P}, \circ \rangle$ by any NCI-likelihood assignment.

Proof. Simply run the construction above with a total order whose cardinality is strictly greater than that of $|\mathbb{P}|$. By construction, any prospective likelihood assignment must assign two sets $A, B \in \mathcal{F}_{\mathcal{L}}$ to the same value $p \in \mathbb{P}$ despite $A \prec B$. $\square$

B.2 Proofs for Section 3.3

Proposition B.6

Let Ω be a finite set of possibilities, $\mathcal{F}$ an algebra over Ω, S_B the 2-Brier score, and $\mathcal{E}$ a tuple of events $\langle E_1, \ldots, E_n \rangle$ with $n \geq 1$ and $E_1, \ldots, E_n \in \mathcal{F}$.

(i) For any 2-forecast f over $\mathcal{E}$, if f is not 2-probabilistic, then there exists a 2-probabilistic 2-forecast f^* such that $S_B(f^*, w) < S_B(f, w)$ for every $w \in \Omega$.

(ii) For any 2-probabilistic 2-forecast f over $\mathcal{E}$, there exists no 2-forecast f^* such that $S(f^*, w) \leq S(f, w)$ for every $w \in \Omega$ and $S(f^*, w) < S(f, w)$ for some $w \in \Omega$.

Proof. Let $f[1]$ denote the element from $[0, 1]^n$ such that $f_i[1] = f[1]_i$ for every i. Let $f[2]$ denote the element from $[0, 1]^n$ such that $f_i[2] = f[2]_i$ for every i.

(i): Suppose f is not 2-probabilistic. It follows immediately that either $f[1]$ is not probabilistic, $f[2]$ is not probabilistic, or both $f[1]$ and $f[2]$ are not probabilistic. In every case, part (i) of the de Finetti (1979); Savage (1971) theorem provides strictly dominating probabilistic forecasts, and so we replace the non-probabilistic forecasts with these probabilistic ones. In the third case for example, part (i) of the de Finetti (1979); Savage (1971) theorem provides a probabilistic forecast $f^*[1]$ and $f^*[2]$ each of which strictly dominate their non-probabilistic counterparts. $f^*_i = \langle f^*[1]_i, f^*[2]_i \rangle$ is then a 2-forecast. Since the 2-Brier score is simple a pair of Brier scores over the first and second components, f^*'s 2-Brier score strictly dominates f's simply in virtue of $f^*[1]$ strictly dominating $f[1]$.

(ii): Suppose that f is a 2-probabilistic 2-forecast. Assume for reductio that there exists a 2-forecast f which weakly dominates it. By definition, the 2-Brier score of f is at least as small as that of f^* in every world and strictly smaller in at least one. By part (ii) of the de Finetti (1979); Savage (1971) theorem, no forecast f' weakly dominates $f^*[1]$, so the first component of f's 2-Brier score cannot be both at least as small as that of f^* in every world and strictly smaller in some world. f must, then, weakly dominate f^* by matching the first component of f^*'s 2-Brier score in every world and doing better in the second component. Applying part (ii) of the de Finetti (1979); Savage (1971) theorem again, however, no forecast f'' weakly dominates $f^*[2]$. The second component of f's 2-Brier score cannot then be both at least as small as that of f^* in every world and strictly smaller in some world. f does not, then, weakly dominate f^*. $\square$

Appendix C

Proofs for Chapter 4

C.1 Proofs for Section 4.3

Proposition C.1

For a propositional language L, consistent $\Sigma \subseteq L$*, and* $\varphi, \psi \in L$*,*

$$\varphi \precsim_{\Sigma}^{PL} \psi \Leftrightarrow [\![\varphi]\!]_{\Sigma} \subseteq [\![\psi]\!]_{\Sigma}.$$

Proof.
($\Rightarrow$)
Suppose that $\varphi \precsim_{\Sigma}^{\mathrm{PL}} \psi$. By definition, it follows that $\Sigma, \varphi \models_{\mathrm{PL}} \psi$, and thus that every Σ-model which makes φ true also makes ψ true, i.e. $[\![\varphi]\!]_{\Sigma} \subseteq [\![\psi]\!]_{\Sigma}$.
($\Leftarrow$)
Suppose that $[\![\varphi]\!]_{\Sigma} \subseteq [\![\psi]\!]_{\Sigma}$, i.e. every Σ-model which makes φ true also makes ψ true. It follows immediately that $\Sigma, \varphi \models_{\mathrm{PL}} \psi$, and thus $\varphi \precsim_{\Sigma}^{\mathrm{PL}} \psi$. $\square$

Proposition C.2

For a propositional language L, consistent $\Sigma \subseteq L$, and $\varphi, \psi \in L$,

(i) Non-triviality
$\bot \prec^{PL}_{\Sigma} \top$.

(ii) Transitivity
For any $\varphi_1, \varphi_2, \varphi_3 \in L$, if $\varphi_1 \precsim^{PL}_{\Sigma} \varphi_2$ and $\varphi_2 \precsim^{PL}_{\Sigma} \varphi_3$, then $\varphi_1 \precsim^{PL}_{\Sigma} \varphi_3$.

(iii) Reflexivity
For any $\varphi \in L$, $\varphi \precsim^{PL}_{\Sigma} \varphi$.

(iv) Boundedness
For any $\varphi \in L$ such that $\varphi \nsim^{PL}_{\Sigma} \bot$ and $\varphi \nsim_{\Sigma} \top$, $\bot \prec^{PL}_{\Sigma} \varphi \prec^{PL}_{\Sigma} \top$.

(v) Monotonicity
If $\varphi \wedge \gamma \sim^{PL}_{\Sigma} \psi \wedge \gamma \sim^{PL}_{\Sigma} \bot$, then $\varphi \precsim^{PL}_{\Sigma} \psi$ if and only if $\varphi \vee \gamma \precsim^{PL}_{\Sigma} \psi \vee \gamma$.

Proof. Characteristics (i)-(v) all follow immediately from the characterization in terms of sets of models. □

Proposition C.3

For a first-order language L, consistent $\Sigma \subseteq L$, and $\varphi, \psi \in L$,

(i) Non-triviality
$\bot \prec^{FOL}_{\Sigma} \top$.

(ii) Transitivity
For any $\varphi_1, \varphi_2, \varphi_3 \in L$, if $\varphi_1 \precsim^{FOL}_{\Sigma} \varphi_2$ and $\varphi_2 \precsim^{FOL}_{\Sigma} \varphi_3$, then $\varphi_1 \precsim^{FOL}_{\Sigma} \varphi_3$.

(iii) Reflexivity
For any $\varphi \in L$, $\varphi \precsim^{FOL}_{\Sigma} \varphi$.

(iv) Boundedness
For any $\varphi \in L$ such that $\varphi \nsim^{FOL}_{\Sigma} \bot$ and $\varphi \nsim^{FOL}_{\Sigma} \top$, $\bot \prec^{FOL}_{\Sigma} \varphi \prec^{FOL}_{\Sigma} \top$.

(v) Monotonicity
If $\varphi \wedge \gamma \sim^{FOL}_{\Sigma} \psi \wedge \gamma \sim^{FOL}_{\Sigma} \bot$, then $\varphi \precsim^{FOL}_{\Sigma} \psi$ if and only if $\varphi \vee \gamma \precsim^{FOL}_{\Sigma} \psi \vee \gamma$.

Proof. (i) Since Σ is consistent, $\Sigma, \bot \models_{\text{FOL}} \top$ and $\Sigma, \top \not\models_{\text{FOL}} \bot$.

(ii) Suppose $\varphi_1 \precsim^{\text{FOL}}_{\Sigma} \varphi_2$ and $\varphi_2 \precsim^{\text{FOL}}_{\Sigma} \varphi_3$. By definition, $\Sigma, \varphi_1 \models_{\text{FOL}} \varphi_2$ and $\Sigma, \varphi_2 \models_{\text{FOL}} \varphi_3$. It follows by elementary metalogic that $\Sigma, \varphi_1 \models_{\text{FOL}} \varphi_3$, and thus that $\varphi_1 \precsim^{\text{FOL}}_{\Sigma} \varphi_3$.

(iii) For any φ, $\Sigma, \varphi \models_{\text{FOL}} \varphi$.

(iv) Suppose that $\varphi \nsim^{\text{FOL}}_{\Sigma} \bot$ and $\varphi \nsim^{\text{FOL}}_{\Sigma} \top$. Since $\Sigma, \bot \models_{\text{FOL}} \varphi$ and $\varphi \nsim^{\text{FOL}}_{\Sigma} \bot$, it must be that $\Sigma, \varphi \not\models_{\text{FOL}} \bot$. It follows that $\bot \prec^{\text{FOL}}_{\Sigma} \varphi$. Similarly, since $\Sigma, \varphi \models_{\text{FOL}} \top$ and $\varphi \nsim^{\text{FOL}}_{\Sigma} \top$, it must be that $\Sigma, \top \not\models_{\text{FOL}} \varphi$. It follows immediately that $\varphi \prec^{\text{FOL}}_{\Sigma} \top$.

(v) Suppose $\varphi \wedge \gamma \sim^{\text{FOL}}_{\Sigma} \psi \wedge \gamma \sim^{\text{FOL}}_{\Sigma} \bot$.
($\Rightarrow$)
Assume that $\varphi \precsim^{\text{FOL}}_{\Sigma} \psi$. By definition, $\Sigma, \varphi \models_{\text{FOL}} \psi$. It follows by elementary metalogic that $\Sigma, \varphi \vee \gamma \models_{\text{FOL}} \psi \vee \gamma$.
($\Leftarrow$)
Assume that $\varphi \vee \gamma \precsim^{\text{FOL}}_{\Sigma} \psi \vee \gamma$. By definition, $\Sigma, \varphi \vee \gamma \models_{\text{FOL}} \psi \vee \gamma$. Since there are no Σ-models which make $\varphi \wedge \gamma$ true, it follows that every Σ-model which makes φ true also makes ψ true, i.e. $\Sigma, \varphi \models_{\text{FOL}} \psi$.

□

C.2 Proofs for Section 4.4

Proposition C.4

Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then, $\mu^{\mathcal{L}}_{\Sigma}$ satisfies Kolmogorov's probability axioms over $\mathcal{F}^{+}_{\mathcal{L}}$.

Proof. First, observe that the definition of $\mu^{\mathcal{L}}_{\Sigma}$ entails that $\mu^{\mathcal{L}}_{\Sigma}$ only assigns rational values between 0 and 1. (K1) thus follows directly from the definition provided. Similarly, (K2) follows immediately from the fact that $\mu^{\mathcal{L}}_{\Sigma}(\Omega_{\mathcal{L}}) = \frac{|\Omega_{\mathcal{L}} \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|} = 1$. Finally, let $E_1, E_2, \ldots$ be a countable sequence of disjoint sets. Given $\Omega_{\mathcal{L}}$ is finite, $E_1, E_2, \ldots$ must produce a finite sequence $E'_1, \ldots, E'_n$ when $\emptyset$ is eliminated.

$$\begin{aligned}
\mu^{\mathcal{L}}_{\Sigma}(\cup^{n}_{i=1} E'_i) &= \frac{|\Omega_{\mathcal{L},\Sigma} \cap \cup^{n}_{i=1} E'_i|}{|\Omega_{\mathcal{L},\Sigma}|} \\
&= \frac{|\Omega_{\mathcal{L},\Sigma} \cap E'_1| + \cdots + |\Omega_{\mathcal{L},\Sigma} \cap E'_n|}{|\Omega_{\mathcal{L},\Sigma}|} \\
&= \frac{|\Omega_{\mathcal{L},\Sigma} \cap E'_1|}{|\Omega_{\mathcal{L},\Sigma}|} + \cdots + \frac{|\Omega_{\mathcal{L},\Sigma} \cap E'_n|}{|\Omega_{\mathcal{L},\Sigma}|} \\
&= \Sigma^{n}_{i=1} \mu^{\mathcal{L}}_{\Sigma}(E'_i).
\end{aligned}$$

Since $\mu^{\mathcal{L}}_{\Sigma}(\emptyset) = 0$, the original $E_1, E_2, \ldots$ sequence also satisfies (K3). ◻

Proposition C.5

Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set, and φ a sentence consistent with $\Omega_{\mathcal{L},\Sigma}$. Then, for any ψ,

$$\mu^{\mathcal{L}}_{\Sigma}(\psi|\varphi) = \mu^{\mathcal{L}}_{\Sigma\cup\{\varphi\}}(\psi)$$

Proof.

$$\begin{aligned}
\mu^{\mathcal{L}}_{\Sigma}(\psi|\varphi) &= \frac{\mu^{\mathcal{L}}_{\Sigma}(\psi \wedge \varphi)}{\mu^{\mathcal{L}}_{\Sigma}(\varphi)} \\
&= \left(\frac{|[\![\psi \wedge \varphi]\!]_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}\right)\left(\frac{|\Omega_{\mathcal{L},\Sigma}|}{|[\![\varphi]\!]_{\mathcal{L},\Sigma}|}\right) \\
&= \frac{|[\![\psi \wedge \varphi]\!]_{\mathcal{L},\Sigma}|}{|[\![\varphi]\!]_{\mathcal{L},\Sigma}|} \\
&= \frac{|[\![\psi]\!]_{\mathcal{L},\Sigma\cup\{\varphi\}}|}{|\Omega_{\mathcal{L},\Sigma\cup\{\varphi\}}|} \\
&= \mu^{\mathcal{L}}_{\Sigma\{\varphi\}}(\psi).
\end{aligned}$$

◻

Proposition C.6

Let L be a propositional language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then,

$$\varphi \precsim^{PL}_{\Sigma} \psi \Rightarrow \varphi \precsim^{\mathcal{L}}_{\Sigma} \psi$$

$$\varphi \precsim^{\mathcal{L}}_{\Sigma} \psi \not\Rightarrow \varphi \precsim^{PL}_{\Sigma} \psi.$$

Proof. Suppose that $\varphi \precsim^{\mathrm{PL}}_{\Sigma} \psi$. By construction, every model of φ is also a model of ψ. It follows immediately that $[\![\varphi]\!]_{\mathcal{L},\Sigma} \subseteq [\![\psi]\!]_{\mathcal{L},\Sigma}$ as well and thus $\varphi \precsim^{\mathcal{L}}_{\Sigma} \psi$.

Consider the propositional language L generated from signature $\sigma = \{P, Q\}$, $\Omega_{\mathcal{L}}$ as every propositional L-model, and $\Sigma = \emptyset$. By definition,

$$P \wedge Q \precsim^{\mathcal{L}}_{\Sigma} P \wedge \neg Q$$

$$P \wedge Q \not\precsim^{\mathrm{PL}}_{\Sigma} P \wedge \neg Q$$

□

C.3 Proofs for Section 4.5

Proposition C.7

For any interpreted propositional languages $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$ and $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+} \rangle$, if $\mathcal{L}^+$ is a completely restricted extension of $\mathcal{L}$, then there exists a bijection $\delta : \Omega_{\mathcal{L}} \to \Omega_{\mathcal{L}^+}$ which preserves the truth of all L-sentences.

Proof. Let $v \in \Omega_{\mathcal{L}}$. By $\mathcal{L}^+$ a completely restricted extension, for every propositional letter P added between L and L^+, there exists a set $R_P \subseteq \Omega_{\mathcal{L}}$ which indicates the truth or falsity of P. Thus, for each propositional letter P, if $v \in R_P$, then all valuations in $\Omega_{\mathcal{L}^+}$ extending v assign P the value T; similarly, for each propositional letter P, if $v \notin R_P$, then all valuations in $\Omega_{\mathcal{L}^+}$ extending v assign P the value F. Put together, these conditions guarantee that there exists a unique valuation $v^+ \in \Omega_{\mathcal{L}^+}$ extending v. Set $\delta(v) = v^+$. Since v^+ is an extension of v, $v^+(\varphi) = v(\varphi)$ for all $\varphi \in L$. □

Proposition C.8

Let interpreted propositional languages $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$ and $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+} \rangle$ as well as a consistent (with respect to $\Omega_{\mathcal{L}}$) theory Σ be given. If $\mathcal{L}^+$ is a completely restricted extension of $\mathcal{L}$ and $\Omega_{\mathcal{L}^+,\Sigma}$ is finite and non-empty, then for any set $A \in \mathcal{F}^+_{\mathcal{L}}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{v^+ \in \Omega_{\mathcal{L}^+} \ : \ v^+|_L \in A\}$.

Proof. Since the truth value of any sentence $\varphi \in \mathcal{L}$ is fixed across the bijection produced by the previous proposition, $\delta[A] = A^+$, $\delta[\Omega_{\mathcal{L},\Sigma}] = \Omega_{\mathcal{L}^+,\Sigma}$, and $\delta[A \cap \Omega_{\mathcal{L},\Sigma}] = A^+ \cap \Omega_{\mathcal{L},\Sigma}$. As a result,

$$\frac{|A \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|} = \frac{|A^+ \cap \Omega_{\mathcal{L}^+,\Sigma}|}{|\Omega_{\mathcal{L}^+,\Sigma}|}$$

and so by definition

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+).$$

□

Proposition C.9

Let $\mathcal{L} = \langle L, \Omega_{\mathcal{L}} \rangle$ be an interpreted propositional language and $\Sigma \subseteq L$ a theory such that $\Omega_{\mathcal{L},\Sigma}$ is finite and nonempty. If $\mathcal{L}^+ = \langle L, \Omega_{\mathcal{L}^+} \rangle$ is a finite, unrestricted extension of $\mathcal{L}$, then for any $A \in \mathcal{F}^+_{\mathcal{L},\Sigma}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{v^+ \in \Omega_{\mathcal{L}^+} \ : \ v^+|_L \in A\}$.

Proof. Let $v_0 \in \Omega_{\mathcal{L},\Sigma}$. We show by induction that, for any extension of n propositional letters, there exists exactly 2^n total valuations in $\Omega_{\mathcal{L}^+,\Sigma}$ which extend v_0.

If $n = 0$, then the result holds trivially.

Suppose that for any expansion of n propositional letters, there exists exactly 2^n total valuations in $\Omega_{\mathcal{L}^+,\Sigma}$ which extend v_0. Faced with an extension of $n+1$ propositional letters, we may rewrite this into an n-letter extension to an intermediary language $\mathcal{L}^n$ (defined as above) followed by a one letter extension to $\mathcal{L}^+$. By hypothesis, we have that there exists exactly 2^n total valuations in $\Omega_{\mathcal{L}^n,\Sigma}$ which extend v_0. Consider now the one letter extension from $\mathcal{L}^n$ to $\mathcal{L}^+$. In particular, let v_n be one of the 2^n extensions of v_0. Since $\Omega_{\mathcal{L}^+,\Sigma} = \{L^+\text{-valuations } v \ : \ v|_L \in \Omega_{\mathcal{L},\Sigma}\} = \{L^+\text{-valuations } v \ : \ v \downarrow_{L^n} \in \Omega_{\mathcal{L}^n,\Sigma}\}$, there exists exactly two valuations in $\Omega_{\mathcal{L}^+,\Sigma}$ which extend v_n; one which maps the new propositional latter to T and one which maps the new propositional letter to F. Since v_n was arbitrary, this holds for every one of the 2^n extensions of v_0; further, none of these extensions can overlap since each of the v_n are distinct. It must be then that there exists exactly $2 * 2^n = 2^{n+1}$ expansions of v_0 in $\Omega_{\mathcal{L}^+,\Sigma}$.

Let $A \in \mathcal{F}^+_{\mathcal{L},\Sigma}$. Then, for every $v_0 \in A$ there exists 2^n distinct extensions of v_0 in $\Omega_{\mathcal{L}^+,\Sigma}$. That is, $|A^+| = 2^n(|A|)$. Similarly, $|\Omega_{\mathcal{L}^+,\Sigma}| = 2^n(|\Omega_{\mathcal{L},\Sigma}|)$. It follows that

$$\frac{|A^+|}{|\Omega_{\mathcal{L}^+,\Sigma}|} = \frac{2^n(|A|)}{2^n(|\Omega_{\mathcal{L},\Sigma}|)} = \frac{|A|}{|\Omega_{\mathcal{L},\Sigma}|}$$

and thus

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{L^+}_{\Sigma}(A^+).$$

□

C.4 Proofs for Section 4.6

Definition C.1 A *first-order signature* σ is a set of symbols each of which is either a constant, a relation symbol with a fixed arity $n \in \mathbb{N} - \{0\}$, or a function symbol with a fixed arity $n \in \mathbb{N} - \{0\}$.

Definition C.2 The *terms* of a first-order signature σ are the strings of symbols defined as follows:

- Every variable is a term of σ.
- Every constant in σ is a term of σ.
- For $n > 0$, if f is an n-ary function symbol from σ and $t_1, \ldots, t_n$ are terms of σ, then $f(t_1, \ldots, t_n)$ is a term of σ as well.
- Nothing else is a term of σ.

Definition C.3 A *closed term* is a term with no variables.

If a term is introduced as $t(\bar{x})$ this indicates that $\bar{x}$ is a sequence $(x_0, x_1, \ldots, x_n)$ of variables, and every variable which occurs in t is among the variables in $\bar{x}$.

Definition C.4 The *atomic formulas* of a signature σ are the strings of symbols generated by:

- If s and t are terms of σ, then '$s = t$' is an atomic formula of σ.
- For $n > 0$, if R is an n-ary relation symbol from σ, and $t_1, \ldots, t_n$ are terms of σ, then the expression $R(t_1, \ldots, t_n)$ is an atomic formula of σ.

Definition C.5 An *atomic sentence* is an atomic formula in which there are no variables or–equivalently–all of whose terms are closed.

Just as with terms, if we introduce an atomic formula φ as $\varphi(\bar{x})$, then $\varphi(\bar{s})$ means the atomic formula got from φ by putting terms from the sequence $\bar{s}$ in place of all occurrences of the corresponding variables in $\bar{x}$.

Definition C.6 The set of well-formed *formulas* from signature σ is the set defined by:

- Every atomic formula from σ is a well-formed formula from σ.
- For any well-formed formulas $\varphi(\bar{x})$ and $\psi(\bar{x})$ from σ,

  $\neg\varphi(\bar{x})$ is a well-formed formula of σ.

  $(\varphi(\bar{x}) \wedge \psi(\bar{x}))$ is a well-formed formula of σ.

  $(\varphi(\bar{x}) \vee \psi(\bar{x}))$ is a well-formed formula of σ.

  $(\varphi(\bar{x}) \rightarrow \psi(\bar{x}))$ is a well-formed formula of σ.

  $(\varphi(\bar{x}) \leftrightarrow \psi(\bar{x}))$ is a well-formed formula of σ.

  $\forall v\ \varphi(\bar{x})$ where v is any variable is a well-formed formula of σ.

  $\exists v\ \varphi(\bar{x})$ where v is any variable is a well-formed formula of σ.

Definition C.7 A variable v in a formula $\varphi(\bar{x})$ of a signature σ is *free* if and only if either

- $\varphi(\bar{x})$ is an atomic formula from signature σ and v occurs in $\varphi(\bar{x})$ or
- $\varphi(\bar{x})$ is a formula from signature σ, v was free up to the last step in the construction of $\varphi(\bar{x})$, and the last step in the construction of $\varphi(\bar{x})$ did not prepend either $\forall v$ or $\exists v$.

Definition C.8 A *first-order language L* with signature σ is the set of formulas from σ with no free variables, also called sentences.

Definition C.9 Let L be a first-order language generated from signature σ_L, and A a canonical-domain model for L. Furthermore, let $t(\bar{x})$ be a term of L and $\bar{a}$ a sequence of elements from A at least as long as $\bar{x}$. Then, by recursion on the complexity of the term t,

- If $t(\bar{x})$ is c for some constant c of the signature L, then $t^A[\bar{a}]$ is c^A.
- If $t(\bar{x})$ is x_i for some variable x_i in $\bar{x}$, then $t^A[\bar{a}]$ is a_i; that is, the corresponding entry in the tuple $\bar{a}$.
- If $t(\bar{x})$ is $f(\bar{s})$ for some function symbol f and tuple of terms $\bar{s}$, then $t^A[\bar{a}]$ is $f^A(s_1^A[\bar{a}], \ldots, s_n^A[\bar{a}])$.

Note that these three clauses fix the denotation of any term in L when it is paired with a tuple of elements from dom(A) of appropriate length. This done, we may now define *truth in a (canonical-domain) model* or the $\Vdash$ relation. Again, the definition is recursive, but now on the complexity of the formula. First, suppose that $\varphi(\bar{x})$ is atomic. Then, given a tuple of elements $\bar{a}$ at least as long as $\bar{x}$,

- $A \Vdash R(t_1, \ldots, t_n)[\bar{a}]$ is true for some relation symbol R in L if and only if $\langle t_1^A[\bar{a}], \ldots, t_n^A[\bar{a}]\rangle \in R^A$, and
- $A \Vdash t = s\ [\bar{a}]$ if and only if $t^A[\bar{a}] = s^A[\bar{a}]$.

Finally, the recursive step:

- $A \Vdash \neg\varphi[\bar{a}]$ if and only if $A \nVdash \varphi[\bar{a}]$
- $A \Vdash \varphi[\bar{a}] \wedge \psi[\bar{a}]$ if and only if $A \Vdash \varphi[\bar{a}]$ and $A \Vdash \psi[\bar{a}]$.
- $A \Vdash \varphi[\bar{a}] \vee \psi[\bar{a}]$ if and only if $A \Vdash \varphi[\bar{a}]$ or $A \Vdash \psi[\bar{a}]$.
- $A \Vdash \forall y\ \varphi[y, \bar{a}]$ if and only if for every element b of A, $A \Vdash \varphi[b, \bar{a}]$
- $A \Vdash \exists y\ \varphi[y, \bar{a}]$ if and only if for some element b of A, $A \Vdash \varphi[b, \bar{a}]$.

Proposition C.10

Let L be a first-order language, $\Gamma \subseteq L$, and $\varphi \in L$. Then,

$$\Gamma \models \varphi \text{ if and only if } \Gamma \models_C \varphi.$$

Proof. Since every canonical-domain model is a first-order model, it suffices to show that for any first-order model $\mathcal{M}$ there exists an isomorphic canonical-domain model $\mathcal{M}'$. Let

$\mathcal{M}$ be given. By definition, there exists a canonical domain with the same cardinality as $|\mathcal{M}|$; let this be the domain of $\mathcal{M}'$. Since $|\mathcal{M}|$ and $|\mathcal{M}'|$ have the same cardinality, there exists a bijection δ between them. For any symbol s from σ_L, set the extension of s as $\delta(s^{\mathcal{M}})$. It follows immediately that δ is an isomorphism between $\mathcal{M}$ and $\mathcal{M}'$ as desired. □

Proposition C.11

Let L be a first-order language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a set of sentences consistent with $\Omega_{\mathcal{L}}$. Then, $\mu^{\mathcal{L}}_{\Sigma}$ satisfies Kolmogorov's probability axioms over $\mathcal{F}^{+}_{\mathcal{L}}$.

Proof. First, observe that the definition of $\mu^{\mathcal{L}}_{\Sigma}$ entails that $\mu^{\mathcal{L}}_{\Sigma}$ only assigns rational values between 0 and 1. (K1) thus follows directly from the definition provided. Similarly, (K2) follows immediately from the fact that $\mu^{\mathcal{L}}_{\Sigma}(\Omega_{\mathcal{L}}) = \frac{|\Omega_{\mathcal{L}} \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|} = 1$. Finally, let $E_1, E_2, \ldots$ be a countable sequence of disjoint sets. Given $\Omega_{\mathcal{L}}$ is finite, $E_1, E_2, \ldots$ must produce a finite sequence $E'_1, \ldots, E'_n$ when $\emptyset$ is eliminated.

$$\begin{aligned}
\mu^{\mathcal{L}}_{\Sigma}(\cup^n_{i=1} E'_i) &= \frac{|\Omega_{\mathcal{L},\Sigma} \cap \cup^n_{i=1} E'_i|}{|\Omega_{\mathcal{L},\Sigma}|} \\
&= \frac{|\Omega_{\mathcal{L},\Sigma} \cap E'_1| + \cdots + |\Omega_{\mathcal{L},\Sigma} \cap E'_n|}{|\Omega_{\mathcal{L},\Sigma}|} \\
&= \frac{|\Omega_{\mathcal{L},\Sigma} \cap E'_1|}{|\Omega_{\mathcal{L},\Sigma}|} + \cdots + \frac{|\Omega_{\mathcal{L},\Sigma} \cap E'_n|}{|\Omega_{\mathcal{L},\Sigma}|} \\
&= \Sigma^n_{i=1} \mu^{\mathcal{L}}_{\Sigma}(E'_i).
\end{aligned}$$

Since $\mu^{\mathcal{L}}_{\Sigma}(\emptyset) = 0$, the original $E_1, E_2, \ldots$ sequence also satisfies (K3). □

Corollary C.1

Let L be a first-order language, $\Omega_{\mathcal{L}}$ a finite model space for L, and $\Sigma \subseteq L$ a set of sentences consistent with $\Omega_{\mathcal{L}}$. Then, for any φ consistent with $\Omega_{\mathcal{L},\Sigma}$ and any ψ,

$$\mu^{\mathcal{L}}_{\Sigma}(\psi|\varphi) = \mu^{\mathcal{L}}_{\Sigma \cup \{\varphi\}}(\psi).$$

Proof.

$$\mu^{\mathcal{L}}_{\Sigma}(\psi|\varphi) = \frac{\mu^{\mathcal{L}}_{\Sigma}(\psi \wedge \varphi)}{\mu^{\mathcal{L}}_{\Sigma}(\varphi)}$$

$$
\begin{aligned}
&= \left(\frac{|[\![\psi \wedge \varphi]\!]_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}\right)\left(\frac{|\Omega_{\mathcal{L},\Sigma}|}{|[\![\varphi]\!]_{\mathcal{L},\Sigma}|}\right) \\
&= \frac{|[\![\psi \wedge \varphi]\!]_{\mathcal{L},\Sigma}|}{|[\![\varphi]\!]_{\mathcal{L},\Sigma}|} \\
&= \frac{|[\![\psi]\!]_{\mathcal{L},\Sigma\cup\{\varphi\}}|}{|\Omega_{\mathcal{L},\Sigma\cup\{\varphi\}}|} \\
&= \mu^{\mathcal{L}}_{\Sigma\cup\{\varphi\}}(\psi).
\end{aligned}
$$

□

Proposition C.12

Let $\mathcal{L}$ be an interpreted first-order language, $\mathcal{L}^+$ a completely restricted extension of $\mathcal{L}$, and $\Sigma \subseteq L$. If $\Omega_{\mathcal{L},\Sigma}$ is non-empty and finite, then for any $A \in \mathcal{F}^+_{\mathcal{L}}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{\mathcal{M}^+ \in \Omega_{\mathcal{L}^+} \ : \ \mathcal{M}^+|_L \in A\}$.

Proof. We construct a bijection δ from $\Omega_{\mathcal{L}}$ to $\Omega_{\mathcal{L}^+}$ which preserves the truth of all L-sentences. Let $\mathcal{M} \in \Omega_{\mathcal{L}}$. Since $\mathcal{L}^+$ is a completely restricted extension, there exists a unique L^+-model $\mathcal{M}^+ \in \Omega_{\mathcal{L}^+}$ such that $\mathcal{M}^+|_L = \mathcal{M}$. Set $\delta(\mathcal{M}) = \mathcal{M}^+$, and note that this construction guarantees that δ is injective. Surjectivity now follows immediately from the fact that $\mathcal{L}^+$ is an extension of $\mathcal{L}$. Note finally that $\mathcal{M} \Vdash \varphi$ if only if $\mathcal{M}^+ \Vdash \varphi$ for any $\varphi \in L$ since the extensions of symbols in L are unchanged between the two. Thus,

$$
\begin{aligned}
\frac{|A \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|} &= \frac{|\delta[A \cap \Omega_{\mathcal{L},\Sigma}]|}{|\delta[\Omega_{\mathcal{L},\Sigma}]|} \\
&= \frac{|A^+ \cap \Omega_{\mathcal{L}^+,\Sigma}|}{|\Omega_{\mathcal{L}^+,\Sigma}|}.
\end{aligned}
$$

□

Proposition C.13

Let $\mathcal{L}$ be a first-order interpreted language, $\mathcal{L}^+$ an unrestricted extension of $\mathcal{L}$ by a finite number of symbols, and $\Sigma \subseteq L$. If every canonical-domain model $\mathcal{M} \in \Omega_{\mathcal{L}}$ has the same domain and $\Omega_{\mathcal{L},\Sigma}$ is both nonempty and finite, then for any set $A \in \mathcal{F}^+_{\mathcal{L}}$,

$$\mu^{\mathcal{L}}_{\Sigma}(A) = \mu^{\mathcal{L}^+}_{\Sigma}(A^+)$$

where $A^+ = \{\mathcal{L}^+\text{-models } \mathcal{M}^+ \in \Omega_{\mathcal{L}^+} \ : \ \mathcal{M}^+|_L \in A\}$.

Proof. We show that $\Omega_{\mathcal{L}^+,\Sigma}$ can be partitioned into parts of equal size with all L^+-models in each part reducing to the same L-model from $\Omega_{\mathcal{L},\Sigma}$. First, note that since $\Omega_{\mathcal{L},\Sigma}$ is finite and $\mathcal{L}^+$ is an extension by a finite number of symbols, $\Omega_{\mathcal{L}^+,\Sigma}$ is likewise finite. Partition the models $\mathcal{M}^+ \in \Omega_{\mathcal{L}^+,\Sigma}$ based on their reduction to L, i.e. based on $\mathcal{M}^+|_L$. Since $\mathcal{L}^+$ is an extension of $\mathcal{L}$, $\mathcal{M}^+|_L \in \Omega_{\mathcal{L}}$, and thus $\mathcal{M}^+|_L \in \Omega_{\mathcal{L},\Sigma}$.

Consider any two parts of this partition. By construction, every model in these parts reduces to $M, M' \in \Omega_{\mathcal{L},\Sigma}$ respectively. Further, these parts must contain every L^+-model which so reduces since $\mathcal{L}^+$ is an unrestricted extension. We now construct a bijection between the parts by pairing models. Let $\mathcal{M}^+$ be a model from the first part which reduces to $\mathcal{M}$. We may fix a unique corresponding model $\mathcal{M}'^+$ from the second part by adjoining $\mathcal{M}^+$'s extension for each new symbol to $\mathcal{M}'$. This process clearly delivers an injective function from the first part to the second. Surjectivity can be established by noting that the process above can be reversed to find a unique model from the first part for any model from the second. The defined function is thus a bijection between these two parts. Since these parts were themselves arbitrary, it follows that every part in the partition has the same cardinality.

Finally, note that each part in the partition assigns identical truth values across L (though different parts may assign different values) since each corresponds to exactly one model from $\Omega_{\mathcal{L},\Sigma}$. Thus, for any $A \in \mathcal{F}_{\mathcal{L}}$,

$$\begin{aligned}\frac{|A \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|} &= \frac{\text{number of parts (in } \Omega_{\mathcal{L}^+,\Sigma}) \text{ which reduce to an } L\text{-model in } A}{\text{total number of parts (in } \Omega_{\mathcal{L},\Sigma})} \\ &= \frac{|A^+ \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L}^+,\Sigma}|}\end{aligned}$$

since all parts are of the same finite size. □

Appendix D

Proofs for Chapter 5

D.1 Proofs for Section 5.1

Proposition D.1

For any finite Ω and $A \subseteq \Omega$, if $g : \Omega \to \Omega'$ is a bijection, then

$$\mu_{Fin}(A, \Omega) = \mu_{Fin}(g[A], g[\Omega]).$$

Proof. Let $g : \Omega \to \Omega'$ be a bijection. By definition, $|A| = |f(A)|$ and $|\Omega| = |\Omega'|$. It follows immediately that $\mu_{\text{Fin}}(A, \Omega) = \mu_{\text{Fin}}(g[A], g[\Omega])$. □

Proposition D.2

Let Ω be a set of possibilities, $\mathcal{F}$ an algebra over Ω, and $\mu : \mathcal{F} \to [0, \infty]$ a measure. Suppose that

(i) Ω *is infinite,*

(ii) $B \in \mathcal{F}$ *is an infinite set with* $0 < \mu(B) < \infty$, *and*

(iii) there exists $A, A', A'_1, A'_2 \in \mathcal{F}$ *such that* $|A| = |A'| = |A'_1| = |A'_2|$, A, A' *partition* B, *and* A'_1, A'_2 *partition* A'.

Then, there exists a bijection $g : \Omega \to \Omega$ such that for some set $A \in \mathcal{F}$,

$$\mu(A) \neq \mu(g[A]).$$

Proof. Since μ is a measure, it must be the case that $\mu(A) + \mu(A') = \mu(B)$. If $\mu(A) \neq \mu(A')$, let g send A to A' bijectively (recall, $|A| = |A'|$) while leaving all other elements of Ω fixed. Otherwise, repeat the decomposition on A' to produce A'_1 and A'_2 with $|A| = |A'| = |A'_1| = |A'_2|$ and $\mu(A'_1) + \mu(A'_2) = \mu(A')$. At least one of $\mu(A'_1)$ and $\mu(A'_2$ must be strictly less than $\mu(A')$. Without loss of generality, suppose that it is $\mu(A'_1)$. Let g send A to A'_1 bijectively while leaving all other elements fixed. In all cases, we have $\mu(A) \neq \mu(g[A])$ as desired. □

D.2 Proofs for Section 5.2

Proposition D.3

Let Ω be given. Then for any choice of $A, B, C \in \mathcal{P}(\Omega)$, $\precsim^\dagger$ satisfies

C0. Nontriviality
$\emptyset \prec \Omega$.

C1a. Reflexivity
$A \precsim A$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

C3a. Nonnegativity
$\bot \precsim A$.

C3b. Bounded
$\bot \precsim A \precsim \Omega$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

Proof.
C0: $\emptyset$ trivially embeds in Ω but not vice-versa.

C1a: $A - A = \emptyset$ for any A, and thus $A \sim^\dagger A$

C2: We proceed by cases.

- Suppose that $A \sim^\dagger B$ and $B \sim^\dagger C$. It follows immediately that $|A - B| = |B - A|$ and $|B - C| = |C - B|$ with all values finite. Rewriting the former equality, we have

$$|A \cap B^c| = |A^c \cap B|$$

The first and second terms here can be expressed as sums:

$$|A \cap B^c| = |A \cap B^c \cap C| + |A \cap B^c \cap C^c|$$
$$|A^c \cap B| = |A^c \cap B \cap C| + |A^c \cap B \cap C^c|.$$

Substituting we thus have

$$|A \cap B^c \cap C| + |A \cap B^c \cap C^c| = |A^c \cap B \cap C| + |A^c \cap B \cap C^c|. \qquad (\star)$$

Rewriting $|B - C| = |C - B|$,

$$|B \cap C^c| = |B^c \cap C|.$$

The first and second terms here too can be expressed as sums:

$$|B \cap C^c| = |A \cap B \cap C^c| + |A^c \cap B \cap C^c|$$
$$|B^c \cap C| = |A \cap B^c \cap C| + |A^c \cap B^c \cap C|.$$

Substituting we thus have

$$|A \cap B \cap C^c| + |A^c \cap B \cap C^c| = |A \cap B^c \cap C| + |A^c \cap B^c \cap C|$$
$$|A^c \cap B \cap C^c| = |A \cap B^c \cap C| + |A^c \cap B^c \cap C| - |A \cap B \cap C^c|.$$

Replacing the $|A^c \cap B \cap C^c|$ term in $(\star)$ by the righthand side above,

$$|A \cap B^c \cap C| + |A \cap B^c \cap C^c| = |A^c \cap B \cap C| + |A \cap B^c \cap C| + |A^c \cap B^c \cap C| - |A \cap B \cap C^c|$$
$$|A \cap B^c \cap C^c| = |A^c \cap B \cap C| + |A^c \cap B^c \cap C| - |A \cap B \cap C^c|$$
$$|A \cap B^c \cap C^c| + |A \cap B \cap C^c| = |A^c \cap B \cap C| + |A^c \cap B^c \cap C|.$$

Collapsing terms,

$$|A - C| = |C - A|.$$

- Suppose that $A \prec^\dagger B$ and $B \prec^\dagger C$. By definition, there exists an injection $g : A - B \to B - A$ and an injection $g' : B - C \to C - B$ but no injection $h : A - B \to B - A$ and no injection $h' : C - B \to B - C$. Represented in terms of disjoint collections and shortening sequences of intersections $A \cap B \cap C$ to ABC,

$$g : AB^cC^c \cup AB^cC \to A^cBC^c \cup A^cBC$$

$$g' : A^cBC^c \cup ABC^c \to A^cB^cC \cup AB^cC.$$

If all collections are finite, then we have by our injections g, g':

$$|AB^cC^c| + |AB^cC| < |A^cBC^c| + |A^cBC|$$
$$|A^cBC^c| + |ABC^c| < |A^cB^cC| + |AB^cC|.$$

Summing

$$|AB^cC^c| + |AB^cC| + |A^cBC^c| + |ABC^c| < |A^cBC^c| + |A^cBC| + |A^cB^cC| + |AB^cC|.$$

Canceling duplicate terms,

$$|AB^cC^c| + |ABC^c| < |A^cBC| + |A^cB^cC|$$

$$|A - C| < |C - A|.$$

It follows by definition that $A - C \prec^{\dagger} C - A$.

Suppose, then, that at least one collection in g, g' above is infinite. Since all collections are disjoint, it follows immediately that there exists an injection

$$g^{+} : AB^cC^c \cup AB^cC \cup A^cBC^c \cup ABC^c \rightarrow A^cBC^c \cup A^cBC \cup A^cB^cC \cup AB^cC$$

An infinite collection must thus appear in the range of g^{+}. Noting further that two of the collections appear in both the domain and range, either A^cBC, A^cB^cC, or both must be both infinite and strictly 'larger' than all the other collections in order to produce the injection discrepancy above (either the axiom of choice or a class-analog is required here depending on the nature of the collections). Exploiting this size discrepancy, it follows immediately that there exist an injection

$$g^{-} : AB^cC^c \cup ABC^c \rightarrow A^cBC \cup A^cB^cC$$

as well as no inverse injection

$$h^{-} : A^cBC \cup A^cB^cC \rightarrow AB^cC^c \cup ABC^c$$

- Suppose that $A \sim^{\dagger} B$ and $B \prec^{\dagger} C$. By definition, $|A - B| = |B - A|$ with both finite and there exists an injection $g' : B - C \rightarrow C - B$ but no injection $h' : C - B \rightarrow B - C$. If all collections are finite, then we have:

$$|AB^cC^c| + |AB^cC| = |A^cBC^c| + |A^cBC|$$
$$|A^cBC^c| + |ABC^c| < |A^cB^cC| + |AB^cC|.$$

Summing

$$|AB^cC^c| + |AB^cC| + |A^cBC^c| + |ABC^c| < |A^cBC^c| + |A^cBC| + |A^cB^cC| + |AB^cC|.$$

Canceling duplicate terms,

$$|AB^cC^c| + |ABC^c| < |A^cBC| + |A^cB^cC|$$
$$|A - C| < |C - A|.$$

It follows by definition that $A - C \prec^{\dagger} C - A$.

Suppose, then, that at least one collection in g' is infinite. Since all collections are disjoint, it follows immediately that there exists an injection

$$g^{+} : AB^cC^c \cup AB^cC \cup A^cBC^c \cup ABC^c \rightarrow A^cBC^c \cup A^cBC \cup A^cB^cC \cup AB^cC.$$

An infinite collection must thus appear in the range of g^{+}. Noting further that two of the collections appear in both the domain and range, A^cB^cC must be both infinite and strictly 'larger' than all the other collections in order to produce the injection discrepancy above (as earlier, either the axiom of choice or a class-analog is required here). Exploiting this size discrepancy, it follows immediately that there exist an injection

$$g^- : AB^cC^c \cup ABC^c \to A^cBC \cup A^cB^cC$$

as well as no inverse injection

$$h^- : A^cBC \cup A^cB^cC \to AB^cC^c \cup ABC^c.$$

By definition, $A - C \prec^\dagger C - A$.

- Suppose that $A \prec^\dagger B$ and $B \sim^\dagger C$. By definition, $|B - C| = |C - B|$ with both finite and there exists an injection $g : A - B \to B - A$ but no injection $h : B - A \to A - B$. If all collections are finite, we have:

$$|AB^cC^c| + |AB^cC| < |A^cBC^c| + |A^cBC|$$
$$|A^cBC^c| + |ABC^c| = |A^cB^cC| + |AB^cC|.$$

Summing

$$|AB^cC^c| + |AB^cC| + |A^cBC^c| + |ABC^c| < |A^cBC^c| + |A^cBC| + |A^cB^cC| + |AB^cC|.$$

Canceling duplicate terms,

$$|AB^cC^c| + |ABC^c| < |A^cBC| + |A^cB^cC|$$
$$|A - C| < |C - A|.$$

It follows immediately that $A - C \prec^\dagger C - A$.

Suppose, then, that at least one collection in g is infinite. Since all collections are disjoint, there exists an injection

$$g^+ : AB^cC^c \cup AB^cC \cup A^cBC^c \cup ABC^c \to A^cBC^c \cup A^cBC \cup A^cB^cC \cup AB^cC$$

An infinite collection must thus appear in the range of g^+. Noting that two of the collections appear in both the domain and range, A^cBC must be both infinite and strictly 'larger' than all the other collections in order to produce the injection discrepancy above (as earlier, either the axiom of choice or a class-analog is required here). Exploiting this size discrepancy, it follows immediately that there exist an injection

$$g^- : AB^cC^c \cup ABC^c \to A^cBC \cup A^cB^cC$$

as well as no inverse injection

$$h^- : A^cBC \cup A^cB^cC \to AB^cC^c \cup ABC^c.$$

By definition, $A - C \prec^{\dagger} C - A$.

C3a: For $A = \emptyset$, $\emptyset \sim^{\dagger} A$ follows from (C1a). Otherwise, $\emptyset \prec^{\dagger} A$ follows immediately by definition.

C3b: Only $A \precsim^{\dagger} \Omega$ remains to be shown. Since $A - \Omega_{\mathcal{L}} = \emptyset$, either $A = \Omega$ and $A \precsim \Omega$ by (C1a) or $A \prec^{\dagger} \Omega$ since only $\emptyset$ embeds in $\emptyset$. $A \precsim \Omega$ follows from each.

C4: For $\Rightarrow$, suppose $A \precsim^{\dagger} B$. Since C disjoint from both A and B, it follows immediately that $(A \cup C) - (B \cup C) = A - B$ and $(B \cup C) - (A \cup C) = B - A$. Since $\prec^{\dagger}$ is completely determined by these two values, $A \cup C \precsim^{\dagger} B \cup C$.

For $\Leftarrow$, suppose $A \cup C \precsim^{\dagger} B \cup C$. Since C disjoint from both A and B, it follows immediately that $(A \cup C) - (B \cup C) = A - B$ and $(B \cup C) - (A \cup C) = B - A$. Since $\prec^{\dagger}$ is completely determined by these two values, $A \precsim^{\dagger} B$.

□

Proposition D.4

Let $\Omega_{\mathcal{L},\Sigma}$ be finite. Then for any choice of $A, B \in \mathcal{P}(\Omega_{\mathcal{L},\Sigma})$, $\precsim^{\dagger}_{\mathcal{L},\Sigma}$ satisfies

C1b. Comparability
$A \precsim B$ or $B \precsim A$.

Proof. Since $\Omega_{\mathcal{L},\Sigma}$ is finite, any $A, B \in \mathcal{P}(\Omega_{\mathcal{L},\Sigma}$ are also finite. It follows immediately that $A - B$ and $B - A$ must be finite, and thus that only case (i) of $\precsim^{\dagger}_{\mathcal{L},\Sigma}$ is ever used. Since the cardinalities of A and B are comparable, so too is $\precsim^{\dagger}_{\mathcal{L},\Sigma}$. □

Proposition D.5

Let $\Omega_{\mathcal{L}}$ be given. Then, for any choice of $A, B, C, C' \in \mathcal{P}(\Omega_{\mathcal{L}})$, $\precsim^*$ satisfies

FE Finite Equivalence
If A, B finite and $|A| = |B|$, then $A \sim B$.

CFE Co-Finite Equivalence
If A, B co-finite and $|A^c| = |B^c|$, then $A \sim B$.

R. Regularity
If $A \neq \emptyset$, then $\emptyset \prec A$.

PW Part-Whole
If $A \subset B$, then $A \prec B$.

FD. Finite Difference
If C, C' finite with $|C| < |C'|$ and $A \cap C = A \cap C' = \emptyset$, then $A \cup C \prec A \cup C'$.

SC. Strong Cardinality
If $|A| < |B|$, then $A \prec B$.

BI. Bijection Invariance
For any bijection $g : \Omega_{\mathcal{L},\Sigma} \to \Omega'$, $A \precsim B$ if and only if $g[A] \precsim g[B]$.

CA. Associativity
If $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, and $G_4 \sim F_3 \cup E_3$ with matching subscripts disjoint, then $C_2 \cup D_2 \sim H_4 \cup G_4$.

Proof.
FE: Follows directly from case 1; subcase 2.

CFE: $A - B \subseteq B^c$ and $B - A \subseteq A^c$. Since A^c, B^c finite, both $A - B$ and $B - A$ are also finite, and we are in case 1 of $\precsim^*$'s definition. Every element c in either A^c or B^c falls into one of three categories: $c \in A^c \cap B^c$, $c \in A^c$ only, or $c \in B^c$ only. Since $|A^c| = |B^c|$, $|A^c - (A^c \cap B^c)| = |B^c - (A^c \cap B^c)|$; that is, the number of category two elements is equal to the number of category three elements since all category one elements appear in both A^c and B^c. Note finally, that $A - B = A \cap B^c = B^c - A^c = B^c - (A^c \cap B^c)$ and $B - A = B \cap A^c = A^c - B^c = A^c - (A^c \cap B^c)$.

R: Follows immediately from PW.

PW: A, B fall into either case 1; subcase 1 or case 3 with the same result.

FD: $A \cup C$ and $A \cup C'$ fall into case 1; subcase 1.

SC: Suppose that $|A| < |B|$. Then, $|A - B| \leq A$ while $|B - A| = |B|$. It follows immediately that $|A - B| < |B - A|$ with the latter infinite. It follows immediately that there exists a one-to-one function $g : A - B \to B - A$ but not vice-versa. By definition, $A \prec^{\dagger} B$.
BI:
$(\Rightarrow)$
Suppose $A \precsim^{\dagger}_{\mathcal{L},\Sigma} B$. If $A \prec^{\dagger}_{\mathcal{L},\Sigma} B$, then there exists a one-to-one function $\delta : A \to B$ but not vice versa. It follows immediately that $g \circ \delta \circ g^{-1}$ is a one-to-one function from $g[A]$ into $g[B]$ and that no one-to-one function from $g[B]$ to $g[A]$ exists. Thus $g[A] \prec^{\dagger}_{\Omega'} g[B]$. If $A \sim^{\dagger}_{\mathcal{L},\Sigma} B$, then $A-B$ and $B-A$ are finite with $|A-B| = |B-A|$. Since $g[A-B] = g[A]-g[B]$ and $g[B - A] = g[B] - g[A]$, it follows immediately that $g[A] - g[B]$ and $g[B] - g[A]$ are also finite with $|g[A] - g[B]| = |g[B] - g[A]|$, and so $g[A] \sim^{\dagger}_{\Omega'} g[B]$.
$(\Leftarrow)$
Suppose that $g[A] \precsim^{\dagger}_{\Omega'} g[B]$. If $g[A] \prec^{\dagger}_{\Omega'} g[B]$, then there exists a one-to-one function $\delta : g[A] \to g[B]$ but not vice-versa. It follows immediately that $g^{-1} \circ \delta \circ g$ is a one-to-one function from A into B and that no one-to-one function from B into A exists. Thus $A \prec^{\dagger}_{\mathcal{L},\Sigma} B$. If $g[A] \sim^{\dagger}_{\Omega'} g[B]$, then $g[A] - g[B]$ and $g[B] - g[A]$ are finite with $|g[A] - g[B]| = |g[B] - g[A]|$. It follows immediately that $A - B$ and $B - A$ are also finite with $|A - B| = |B - A|$, and so $A \sim^{\dagger}_{\mathcal{L},\Sigma} B$.

Suppose that $A_1 \sim H_4$, $B_1 \sim E_3$, $C_2 \sim A_1 \cup B_1$, $D_2 \sim F_3$, and $G_4 \sim F_3 \cup E_3$ with matching subscripts disjoint. Then, there exists C_2^A and C_2^B a partition of C_2 such that $C_2^A \sim A_1$ and $C_2^B \sim B_1$ (simply divide the finite number of elements in $C_2 - (A_1 \cup B_1)$ so that $C_2^A - A_1 = A_1 - C_2$ and $C_2^B - B_1 = B_1 - C_2$). Then, $C_2 \cup D_2 = C_2^A \cup C_2^B \cup D_2$ where $C_2^A \sim H_4$, $C_2^B \sim E_3$, and $D_2 \sim F_3$. By (C4) (in particular, the comparative formulation of Decomp), $C_2^B \cup D_2 \sim E_3 \cup F_3$. By (C2), $C_2^B \cup D_2 \sim G_4$. By (C4) (in particular, the comparative formulation of Decomp), $C_2 \cup D_2 = C_2^A \cup C_2^B \cup D_2 \sim H_4 \cup G_4$ as desired. □

Proposition D.6

Let L be a first-order language and $\Sigma \subseteq L$ a consistent theory. Taking all canonical-domain models for L and any extension of L as $\Omega_{\mathcal{L}}$, for any $\varphi, \psi \in L$,

$$\varphi \precsim^{\dagger}_{\Sigma} \psi \text{ if and only if } \varphi \precsim^{FOL}_{\Sigma} \psi.$$

Proof. $(\Leftarrow)$ is trivial.
$(\Rightarrow)$
Suppose $\varphi \not\precsim^{FOL}_{\Sigma} \psi$. By definition, there exists an L-model $\mathcal{M}_{\varphi} \in [\![\varphi \wedge \neg\psi]\!]_{\Omega_{\mathcal{L},\Sigma}}$. Since

arbitrary extensions of L are allowed, it follows that there exists infinitely many models $\mathcal{M}_\varphi^+$ which reduce to $\mathcal{M}_\varphi$ over L. Every such model is a model of φ but not ψ, and so the difference between φ-models and ψ-models in $\Omega_{\mathcal{L},\Sigma}$ is infinite.

It suffices now to show that there exists an injection from $\psi \wedge \neg\varphi$-models to $\varphi \wedge \neg\psi$-models. If $\psi \precsim_\Sigma^{FOL} \varphi$, this is trivial. If $\psi \not\precsim_\Sigma^{FOL} \varphi$, there exists $\mathcal{M}_\psi$ a $\psi \wedge \neg\varphi$-model in L. Let $\mathcal{M}_\varphi^+$ be a $\varphi \wedge \neg\psi$-model in some extension L^+ of L. Both the domain of $\mathcal{M}_\varphi^+$ and the constants in the signature of L^+ have a fixed cardinality; let $\aleph_i$ be the greater of these. By the axiom of choice, the set of cardinality $\aleph_{i+1}$ from the canonical-domain class underlying these models admits of a well-ordering; let a_{i+1} be the least element from this set which both does not appear in a set of lesser cardinality in the canonical-domain class and which does not appear in any form as a subscript of a constant in L. Such an element must exist given the definition of $\aleph_i$. Extend L^+ to a language L^{++} by adding a constant $c_{a_{ib}}$ for every element $b \in \mathcal{M}_\varphi^+$, a constant $c_{a_{ic}}$ for every constant c in $L^+ +$, a constant $c_{a_{i\, f(\bar{b})=b'}}$ whenever $f^{\mathcal{M}_\varphi^+}(\bar{b}) = b'$ for a function symbol f in L^+, and a constant $c_{a_{i\, R(\bar{b})}}$ whenever $R^{\mathcal{M}_\varphi^+}(\bar{b})$ for a relation symbol R in L^+. Define a canonical-domain L^{++}-model $\mathcal{M}^*$ by setting all constants not in L to a_0 and all functions/relations not in L to the empty set. Finally, set all symbols from L to their extension in $\mathcal{M}_\psi$. It follows immediately that $\mathcal{M}^* \Vdash \psi \wedge \neg\varphi$. Finally, any $\mathcal{M}_\varphi^+$ must give rise to a unique $\mathcal{M}^*$ since the signature of $\mathcal{M}^*$ must contain a collection of constants which are all subscripted with the same element a_i which does not appear elsewhere and which contains a complete copy of all other constants in the language as a further subscript. Having identified this collection, $\mathcal{M}_\varphi^+$ may be recovered by noting that the c_{a_i} constants encode $\mathcal{M}_\varphi^+$'s atomic diagram. □

Proposition D.7

Let Ω be a set and $\precsim$ a binary relation on $\mathcal{P}(\Omega)$ satisfying

WC. Weak Cardinality
If $|A| < |B|$, then $B \not\precsim A$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

PI. Permutation Invariance
For any permutation $\pi : \Omega \to \Omega$, $A \precsim B$ if and only if $\pi[A] \precsim \pi[B]$.

Then, $A \prec B$ only if $A \prec^\dagger B$.

Proof. Suppose $A \prec B$. By definition, $A \precsim B$ and $B \not\precsim A$. By C4, it follows that $A - B \precsim B - A$ and $B - A \not\precsim A - B$. Suppose for reductio that $A \not\prec^\dagger B$. Thus, either $A \precsim^\dagger B$ and $B \precsim^\dagger A$, $A \not\precsim^\dagger B$ and $B \not\precsim^\dagger A$, or $A \not\precsim^\dagger B$ and $B \precsim^\dagger A$.

Case 1: Suppose that $A \precsim^\dagger B$ and $B \precsim^\dagger A$. Then, $A - B$ and $B - A$ are both finite and $|A - B| = |B - A|$. It follows immediately that there exists a permutation $\pi : \Omega \to \Omega$ which swaps $A-B$ and $B-A$ but leaves all other objects fixed, viz. for a bijection $g : A-B \to B-A$

$$\pi(x) = \begin{cases} g(x) & x \in A - B \\ g^{-1}(x) & x \in B - A \\ x & \text{otherwise.} \end{cases}$$

By the construction of π, $\pi[A - B] \not\precsim \pi[B - A]$, contradicting permutation invariance.

Case 2: Suppose that $A \not\precsim^\dagger B$ and $B \not\precsim^\dagger A$. Then, both $A - B$ and $B - A$ are infinite and $|A| = |B|$. It follows immediately that there again exists a permutation $\pi : \Omega \to \Omega$ which swaps A and B but leaves all other objects fixed. By the construction of π, $\pi[A - B] \not\precsim \pi[B - A]$, contradicting permutation invariance.

Case 3: Suppose $A \not\precsim^\dagger B$ and $B \precsim^\dagger A$. Then, $|B - A| < |A - B|$, contradicting weak cardinality.

By reductio, $A \prec^\dagger B$.

□

Corollary D.1

Let Ω be given and let $\precsim$ be a strict extension of $\precsim^\dagger$. Then, at least one of the following is false:

WC. Weak Cardinality
If $|A| < |B|$, then $B \not\prec A$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

PI. Permutation Invariance
For any permutation $\pi : \Omega \to \Omega$, $A \precsim B$ if and only if $\pi[A] \precsim \pi[B]$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

PW. Part-Whole
If $A \subset B$, then $A \prec B$.

Proof. Suppose $A \not\precsim^\dagger B$ and $B \not\precsim^\dagger A$. By definition, $A - B$ and $B - A$ are both infinite with $|A - B| = |B - A|$. By the result above extending $\precsim^\dagger$ into a binary relation $\precsim$ on $\mathcal{P}(\Omega)$ by setting either $A \prec B$ or $B \prec A$ contradicts either (WC), (C4), or (PI). Suppose, then, that we set $A \sim B$. By C4, $A - B \sim B - A$. Note that since $B - A$ is infinite, it is possible to find $C \subset B - A$ such that $|C| = |B - A|$; it follows immediately that $C \prec^\dagger B - A$ and $C \not\precsim^\dagger A - B$. If we set only $A \sim B$, we now have a failure of transitivity since $C \precsim B - A$, $B - A \precsim A - B$, but $C \not\precsim A - B$.

If we wish to retain transitivity, we must thus have $C \precsim A - B$. Note, however, that there exists a permutation which swaps elements of $A - B$ and C but leaves all other elements fixed, viz. for a bijection $g : A - B \to C$ define

$$\pi(x) = \begin{cases} g(x) & x \in A - B \\ g^{-1}(x) & x \in C \\ x & \text{otherwise.} \end{cases}$$

We must thus also have $A - B \precsim C$, and thus $B - A \sim A - B \sim C$, contradicting PW.

$\square$

Corollary D.2

Let Ω be given and let $\precsim$ be a strict extension of $\precsim^{\dagger}$ satisfying:

WC. Weak Cardinality
If $|A| < |B|$, then $B \not\precsim A$.

C4. Monotonicity
If $A \cap C = B \cap C = \emptyset$, then $A \precsim B$ if and only if $A \cup C \precsim B \cup C$.

PI. Permutation Invariance
For any permutation $\pi : \Omega \to \Omega$, $A \precsim B$ if and only if $\pi[A] \precsim \pi[B]$.

C2. Transitivity
If $A \precsim B$ and $B \precsim C$, then $A \precsim C$.

Then,

(i) $\emptyset \sim C$ *for any* $C \in \mathcal{P}(\Omega)$ *with* $|C| \leq |A - B|$ *and*

(ii) $\Omega \sim C$ *for any* $C \in \mathcal{P}(\Omega)$ *with* $|C^c| \leq |A - B|$.

Proof. Noting that C was arbitrary above, we have $B - A \sim A - B \sim C$ for every $C \subset B - A$ with $|C| = |B - A|$.

Let C_1, C_2 partition $B - A$ such that $|C_1| = |C_2| = |B - A|$. Then, $C_2 = (B - A) - C_1 \sim A - B$ by above. By (C4), $B - A \sim (A - B) \cup C_1$. By (C2), $A - B \sim (A - B) \cup C$. By (C4), $\emptyset \sim C_1$. Since there exists a permutation π which maps C_1 to any subset of Ω with the same cardinality, $\emptyset \sim C$ for any $C \in \mathcal{P}(\Omega)$ with $|C| = |B - A|$. By (WC), it follows immediately that $\emptyset \sim C$ for any $C \in \mathcal{P}(\Omega)$ with $|C| \leq |B - A|$.

Note further that for any $C \in \mathcal{P}(\Omega)$ with $|C^c| \leq |B - A|$, $C^c \sim \emptyset$ and thus by (C4) $\Omega \sim C$. All co-$|A - B|$ sets thus collapse into equivalence with Ω. $\square$

Proposition D.8

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then $\precsim^{\dagger}$ on $\mathcal{F}^{+}_{\mathcal{L},\Sigma}$ is uniquely represented in $\langle \mathbb{Q} \cap [0,1], \leq, + \rangle$ by

$$\mu(E) = \frac{|E|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

Proof. Let L, $\Omega_{\mathcal{L}}$, and Σ be given. Define $\mu : \mathcal{F}^{+}_{\mathcal{L}} \to \mathbb{Q} \cap [0,1]$ by

$$\mu(E) = \frac{|E|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

μ is clearly an NCI-assignment into $\mathbb{Q} \cap [0,1]$. Noting that μ is the first-order atomic confirmation ranking, μ is an NCI-assignment into $\langle \mathbb{Q} \cap [0,1], \leq, + \rangle$. Finally, suppose that $A \precsim^{\dagger} B$. By definition, $A - B$ and $B - A$ are finite with $|A - B| \leq |B - A|$. Since $\Omega_{\mathcal{L},\Sigma}$ is finite, it follows immediately that $|A| \leq |B|$ and thus $\mu(A) \leq \mu(B)$. Suppose now that $\mu(A) \leq \mu(B)$. By definition, $|A| \leq |B|$. Since $\Omega_{\mathcal{L},\Sigma}$ is finite, it follows immediately that $A - B$ and $B - A$ are finite with $|A - B| \leq |B - A|$, i.e. $A \precsim^{\dagger} B$.

For uniqueness, note that $\Omega_{\mathcal{L},\Sigma}$ decomposes into a finite number of equal events. For any $n \in \mathbb{N}^{+}$, there exists exactly one $q \in \mathbb{Q} \cap [0,1]$ which added to itself n times equals 1. Any representing μ for the likelihood structure thus maps these minimal events to the same value. □

Corollary D.3

Let L be a formal language with Boolean connectives, $\Omega_{\mathcal{L}}$ a model space for L, and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then $\precsim^{\dagger}$ on $\mathcal{F}^{+}_{\mathcal{L}}$ is uniquely represented in $\langle \mathbb{Q} \cap [0,1], \leq, + \rangle$ by

$$\mu^{+}(E) = \frac{|E \cap \Omega_{\mathcal{L},\Sigma}|}{|\Omega_{\mathcal{L},\Sigma}|}.$$

Proof. Let L, $\Omega_{\mathcal{L}}$, and Σ be given. Suppose $A, B \in \Omega_{\mathcal{L}}$. By definition,

$$\begin{aligned} A \precsim^{\dagger} B & \text{ if and only if } A \cap \Omega_{\mathcal{L},\Sigma} \precsim^{\dagger} B \cap \Omega_{\mathcal{L},\Sigma} \\ & \text{ if and only if } \mu(A \cap \Omega_{\mathcal{L},\Sigma}) \leq \mu(B \cap \Omega_{\mathcal{L},\Sigma}) \\ & \text{ if and only if } \mu^{+}(A) \leq \mu^{+}(B). \end{aligned}$$

Uniqueness follows immediately from the previous result. □

Proposition D.9

Let $\mathcal{L}^+ = \langle L^+, \Omega_{\mathcal{L}^+}\rangle$ be an n-uniform model space expansion of $\mathcal{L} = \langle L, \Omega_{\mathcal{L}}\rangle$ and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then, there exists a partition $\{P_i, \dots\}$ of $\Omega_{\mathcal{L}^+,\Sigma}$ such that for any i, j:

$$|P_i| = |P_j| = n;$$

For any $\mathcal{M} \in \Omega_{\mathcal{L},\Sigma}$ there exists a unique P_i such that $\mathcal{M}^+ \in P_i \Rightarrow \mathcal{M}_i^+|_L^{L^+} = \mathcal{M}$;

For any P_i there exists a unique $\mathcal{M} \in \Omega_{\mathcal{L},\Sigma}$ such that $\mathcal{M}^+ \in P_i \Rightarrow \mathcal{M}^+|_L^{L^+} = \mathcal{M}$.

Proof. Since every partition corresponds to a model $\mathcal{M}$ in $\Omega_{\mathcal{L}}$, every model in a part agrees on the truth of Σ. Further since Σ is consistent (with respect to $\Omega_{\mathcal{L}}$), at least one part makes Σ true. Eliminating all parts which make Σ false from the *n*-uniform partition thus produces the desired partition over $\Omega_{\mathcal{L}^+,\Sigma}$. □

Proposition D.10

Let $\Omega_{\mathcal{L}^+}$ be an n-uniform model space expansion of $\Omega_{\mathcal{L}}$ and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then for any $E \in \mathcal{F}^+_{\mathcal{L},\Sigma}$,

$$\mu(E) = \mu^+(E^+)$$

where $E^+ = \{\mathcal{M}^+ \in \Omega_{\mathcal{L}^+,\Sigma} \ : \ \mathcal{M}^+|_L^{L^+} \in E\}$ and the functions $\mu : \mathcal{F}^+_{\mathcal{L},\Sigma} \to \mathbb{Q} \cap [0,1]$, $\mu^+ : \mathcal{F}^+_{\mathcal{L}^+,\Sigma} \to \mathbb{Q} \cap [0,1]$ are the unique representing assignments of $\precsim^\dagger$ into $\langle \mathbb{Q} \cap [0,1], \leq, +\rangle$.

Proof.

$$\begin{aligned}\mu(E) &= \frac{|E|}{|\Omega_{\mathcal{L},\Sigma}|}\\ &= \frac{(n)|E|}{(n)|\Omega_{\mathcal{L},\Sigma}|}.\end{aligned}$$

By $\Omega_{\mathcal{L}^+}$ an *n*-uniform expansion,

$$\begin{aligned}&= \frac{|E^+|}{|\Omega_{\mathcal{L}^+,\Sigma}|}\\ &= \mu^+(E^+).\end{aligned}$$

□

Proposition D.11

Let Ω_A and Ω_B be finite but non-empty sets of canonical-domain models with $A \subseteq \Omega_A$ and $B \subseteq \Omega_B$. Then, $\langle A, \Omega_A \rangle \precsim^{PU} \langle B, \Omega_B \rangle$ if and only if $\mu_{\Omega_A}(A) \leq \mu_{\Omega_B}(B)$ where $\mu_{\Omega_A} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ and $\mu_{\Omega_B} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ are the unique representations of $\precsim^{\dagger}_{\Omega_A}$ and $\precsim^{\dagger}_{\Omega_B}$ in $\langle \mathbb{Q} \cap [0,1], + \rangle$.

Proof.
($\Rightarrow$)
Suppose $\langle A, \Omega_A \rangle \precsim^{PU} \langle B, \Omega_B \rangle$. This comparison either derives from (P), from (U), or from closing under transitivity. In the first case, $A \precsim^{\dagger}_{\Omega} B$ and $\Omega_B \precsim^{\dagger}_{\Omega} \Omega_A$. Since all sets are finite, $|A| \leq |B|$ and $|\Omega_B| \leq |\Omega_A|$. It follows immediately that $\frac{|A|}{|\Omega_A|} \leq \frac{|B|}{|\Omega_B|}$ and thus by definition $\mu_A(A) \leq \mu_B(B)$.

In the second case, either Ω_A or Ω_B is the n-uniform extension. If the former, both $|A| = n|B|$ and $|\Omega_A| = (n)|\Omega_B|$. It follows immediately that $\frac{|A|}{|\Omega_A|} \leq \frac{|B|}{|\Omega_B|}$ and thus by definition $\mu_A(A) \leq \mu_B(B)$. If the latter, both $n|A| = |B|$ and $(n)|\Omega_A| = |\Omega_B|$. It follows immediately that $\frac{|A|}{|\Omega_A|} \leq \frac{|B|}{|\Omega_B|}$ and thus by definition $\mu_A(A) \leq \mu_B(B)$.

For the third case, note that that the first two cases have established that the preorder induced by representation in $\langle \mathbb{Q} \cap [0,1], + \rangle$ contains all comparisons required by (P) and (U). Since $\precsim^{PU}$ is the transitive closure of these and the preorder induced by representation in $\langle \mathbb{Q} \cap [0,1], + \rangle$ is transitive, the induced preorder must contain all instances of $\precsim^{PU}$. It follows that $\mu_A(A) \leq \mu_B(B)$.
($\Leftarrow$)
Suppose $\mu_A(A) = \frac{p_A}{q_A} \leq \mu_B(B) = \frac{p_B}{q_B}$. Take a q_B-uniform extension of Ω_A to Ω_A^+ and a q_A-uniform extension of Ω_B to Ω_B^+. A lifts to a set A^+ in Ω_A^+ of cardinality $p_A * q_B$ while B lifts to a set B^+ in Ω_B^+ of cardinality $p_B * q_A$. By assumption, $p_A * q_B \leq p_B * q_A$. It follows immediately that $A^+ \precsim^{\dagger}_{\Omega} B^+$ and $\Omega_B^+ \precsim^{\dagger}_{\Omega} \Omega_A^+$, and thus by (P) we have $\langle A^+, \Omega_A^+ \rangle \precsim^{PU} \langle B^+, \Omega_B^+ \rangle$. By (U) and $A^+ \precsim^{\dagger}_{\Omega_A^+} A^+$, $\langle A, \Omega_A \rangle \precsim^{PU} \langle A^+, \Omega_A^+ \rangle$. By (U) and $B^+ \precsim^{\dagger}_{\Omega_B^+} B^+$, $\langle B^+, \Omega_B^+ \rangle \precsim^{PU} \langle B, \Omega_B \rangle$. Since $\precsim^{PU}$ is the transitive closure, $\langle A, \Omega_A \rangle \precsim^{PU} \langle B, \Omega_B \rangle$. □

Proposition D.12

Let Ω_A be a collection of canonical-domain models and A, A' subcollections of Ω_A. Then,

$$\langle A, \Omega_A \rangle \precsim^{PU} \langle A', \Omega_A \rangle \textit{ if and only if } A \precsim^{\dagger}_{\Omega_A} A'.$$

Proof.
($\Rightarrow$)
Suppose that $\langle A, \Omega_A \rangle \precsim^{PU} \langle A', \Omega_A \rangle$. If Ω_A is finite, then $A \precsim^{\dagger}_{\Omega_A} A'$ follows from the previous proposition. Suppose further, then, that Ω_A is infinite.

An aside on the the exact structure of $\precsim^{PU}$ is useful here. Note first that closing all instances of (U) under transitivity reduces all pairs $\langle B, \Omega_B \rangle$ into equivalence classes where any pair with an infinite Ω_B is alone. The equivalence classes induced by the transitive closure of all (U)-comparisons are further reduced by those $\sim$ relations required by (P), viz. if $B \sim^{\dagger}_{\Omega} C$ and $\Omega_C \sim^{\dagger}_{\Omega} \Omega_B$, then $\langle B, \Omega_B \rangle \sim \langle C, \Omega_C \rangle$. The antecedent here is equivalent to $B - C$, $C - B$, $\Omega_B - \Omega_C$, and $\Omega_C - \Omega_B$ are all finite with $|B - C| = |C - B|$ and $|\Omega_B - \Omega_C| = |\Omega_C - \Omega_B|$. The reduced equivalence classes are thus the finite n-uniform extension classes plus finite replacements in each component. If Ω_B is infinite, equivalent pairs at this stage are exclusively induced by (P).

The only remaining comparisons and the only non-$\sim$ comparisons between equivalence classes are induced by strictly one-sided instances of (P). We show that these are already transitive; whenever $\langle B, \Omega_B \rangle \precsim \langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle \precsim \langle D, \Omega_D \rangle$ by (P) with $\langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle$ in the same equivalence class, there exists $\langle B^*, \Omega_{B^*} \rangle$ and $\langle D^*, \Omega_{D*} \rangle$ equivalent to $\langle B, \Omega_B \rangle$ and $\langle D, \Omega_D \rangle$ respectively such that $\langle B^*, \Omega_{B^*} \rangle \precsim \langle D^*, \Omega_{D^*} \rangle$ by (P). Suppose that $\langle B, \Omega_B \rangle \precsim \langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle \precsim \langle D, \Omega_D \rangle$ by (P) with $\langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle$ equivalent.

- Ω_B, Ω_C, and Ω_D finite. It follows immediately that $\Omega_{C'}$ is finite as well. Further by $\langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle$ in the same equivalence class, $\frac{|C|}{|\Omega_C|} = \frac{|C'|}{|\Omega_{C'}|}$. Thus,

$$\frac{|B|}{|\Omega_B|} \leq \frac{|C|}{|\Omega_C|} = \frac{|C'|}{|\Omega_{C'}|} \leq \frac{|D|}{|\Omega_D|}.$$

  Taking a $|\Omega_D|$-uniform extension of Ω_B and a $|\Omega_B|$-uniform extension of Ω_D, gives a pair $\langle B^*, \Omega_{B^*} \rangle$ and $\langle D^*, \Omega_{D^*} \rangle$ equivalent to $\langle B, \Omega_B \rangle$ and $\langle D, \Omega_D \rangle$ respectively. Further, $|\Omega_{B^*}| = |\Omega_{D^*}|$ and $|B^*| \leq |D^*|$. $\langle B^*, \Omega_{B^*} \rangle \precsim \langle D^*, \Omega_{D^*} \rangle$ follows by (P).

- Ω_B infinite; Ω_C, and Ω_D finite. It follows immediately that $\Omega_{C'}$ is finite as well. Since $|\Omega_B|$ is larger than $|\Omega_{D^*}|$ for any pair $\langle D^*, \Omega_{D^*} \rangle$ equivalent to $\langle D, \Omega_D \rangle$, we need only find an n-uniform extension of Ω_D with $|D^*| \geq |B|$. Since B is finite (Ω_C is finite and $|B| \leq |C|$ by hypothesis), this is always possible (if D is empty, so too is B; otherwise, take a $|B|$-uniform extension). $\langle B, \Omega_B \rangle \precsim \langle D^*, \Omega_{D^*} \rangle$ then follows by (P).

- Ω_B finite; Ω_C infinite. Impossible; $\Omega_C \not\precsim^{\dagger} \Omega_B$.

- Ω_C finite; Ω_D infinite. Impossible; $\Omega_{C'}$ must be finite as well but then $\Omega_D \not\precsim^{\dagger} \Omega_{C'}$.

- Ω_B and Ω_C infinite; Ω_D finite. Since $\langle C', \Omega_{C'} \rangle \precsim \langle D, \Omega_D \rangle$ by (P), $|C'| \leq |D|$ and thus finite. It follows immediately that C is finite as well. Since $\langle B, \Omega_B \rangle \precsim \langle C, \Omega_C \rangle$ by (P), $|B| \leq |C|$ and thus finite. Since $|\Omega_B|$ is larger than $|\Omega_{D^*}|$ for any pair $\langle D^*, \Omega_{D^*} \rangle$ equivalent to $\langle D, \Omega_D \rangle$, we need only find an n-uniform extension of Ω_D with $|D^*| \geq |B|$. Since B is finite, this is always possible (if D is empty, so too is B; otherwise, take a $|B|$-uniform extension). $\langle B, \Omega_B \rangle \precsim \langle D^*, \Omega_{D^*} \rangle$ then follows by (P).

- Ω_B, Ω_C, and Ω_D infinite. $\langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle$ must be equivalent by (P). $B \precsim^{\dagger} D$ and $\Omega_D \precsim^{\dagger} \Omega_B$ follows by the transitivity of $\precsim^{\dagger}$. $\langle B, \Omega_B \rangle \precsim \langle D, \Omega_D \rangle$ then follows by

(P).

Returning to $\langle A, \Omega_A \rangle \precsim^{PU} \langle A', \Omega_A \rangle$ with Ω_A infinite, it must thus be either that $\langle A, \Omega_A \rangle \precsim^{PU} \langle A', \Omega_A \rangle$ is a result of closing under those $\sim$ relations required by (P) or by a strictly one-sided instance of (P). In either case, $A \precsim^{\dagger}_{\Omega} A'$ and thus $A \precsim^{\dagger}_{\Omega_A} A'$ follows immediately.

($\Leftarrow$)
Suppose that $A \precsim^{\dagger}_{\Omega_A} A'$. Then, either $A - A'$ and $A' - A$ are finite and have equal cardinality or there exists an injection $g : A - A' \to A' - A$ but not vice versa. In either case, $A \precsim^{\dagger}_{\Omega} A'$ follows immediately. By (P), we then have $\langle A, \Omega_A \rangle \precsim^{PU} \langle A', \Omega_A \rangle$. □

Proposition D.13

Let $\Omega_{\mathcal{L}^+}$ be an n-uniform model space expansion of $\Omega_{\mathcal{L}}$ for any $n \in \mathbb{N}^+$ and $\Sigma \subseteq L$ a consistent (with respect to $\Omega_{\mathcal{L}}$) set. Then for any A, B from $\mathcal{F}^+_{\mathcal{L},\Sigma}$,

$$A \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B \text{ if and only if } A^+ \precsim^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$$

where A^+ and B^+ are the collections containing all and only those $\Omega_{\mathcal{L}^+,\Sigma}$ models which extend models in A and B respectively.

Proof. If $\Omega_{\mathcal{L},\Sigma}$ is finite, then the result follows from unique representation in $\langle \mathbb{Q} \cap [0, 1], + \rangle$. Suppose, then, that $\Omega_{\mathcal{L},\Sigma}$ is infinite.

($\Rightarrow$)
Suppose that $A \precsim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B$. By definition, either $A \sim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B$ or $A \prec^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B$. In the first case, $|A^+ - B^+| = |B^+ - A^+| = (n)|A - B| = (n)|B - A|$ by $\mathcal{L}^+$ an n-uniform expansion, and thus $A^+ \sim^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$. In the second case, there exists an injection g from $A - B$ to $B - A$ but not vice-versa. Extend g to an injection g^+ from $A^+ - B^+$ to $B^+ - A^+$ by sending the kth $\mathcal{M}^+$ extending $\mathcal{M}^+|_L$ to the kth model extending $g(\mathcal{M}^+|_L)$ (the numbering for any given model can be arbitrary). Finally, no injection from $B^+ - A^+$ to $A^+ - B^+$ is possible since it would immediately produce an injection from $B - A$ to $A - B$.

($\Leftarrow$)
Suppose that $A^+ \precsim^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$. By definition, either $A^+ \sim^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$ or $A^+ \prec^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$. In the first case, $|A^+ - B^+| = |B^+ - A^+| = (n)|A - B| = (n)|B - A|$ by $\mathcal{L}^+$ an n-uniform expansion, and thus $A \sim^{\dagger}_{\Omega_{\mathcal{L},\Sigma}} B$. In the second case, there exists an injection g^+ from $A^+ - B^+$ to $B^+ - A^+$ but not vice-versa. If $A^+ - B^+$ and $B^+ - A^+$ are finite, then $|A^+ - B^+| = (n)|A - B|$ and $|B^+ - A^+| = (n)|B - A|$ by $\mathcal{L}^+$ an n-uniform expansion, and thus $A^+ \prec^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$. If $A^+ - B^+$ is finite and $B^+ - A^+$ infinite, then $|A^+ - B^+| = (n)|A - B|$ by $\mathcal{L}^+$ an n-uniform expansion, and thus $A^+ \prec^{\dagger}_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$. If both $A^+ - B^+$ and $B^+ - A^+$ are infinite, we have by choice (Tarski's

theorem) that $A - B \approx (A-B)^n \approx A^+ - B^+$ and $B - A \approx (B-A)^n \approx B^+ - A^+$. It follows immediately that there exists an injection from $A - B$ to $B - A$ and not vice-versa. Thus, $A^+ \prec^\dagger_{\Omega_{\mathcal{L}^+,\Sigma}} B^+$. □

Proposition D.14

Let Ω be the collection of all atomic possibilities, Ω_A and Ω_B non-empty subcollections, and A, B subcollections of Ω_A and Ω_B, respectively. $\precsim^{PUCI}$ satisfies

C0. Nontriviality
$\langle \emptyset, \Omega \rangle \prec \langle \Omega, \Omega \rangle$.

C1a. Reflexivity
$\langle A, \Omega_A \rangle \precsim \langle A, \Omega_A \rangle$.

C2. Transitivity
If $\langle A, \Omega_A \rangle \precsim \langle B, \Omega_B \rangle$ and $\langle B, \Omega_B \rangle \precsim \langle C, \Omega_C \rangle$, then $\langle A, \Omega_A \rangle \precsim \langle C, \Omega_C \rangle$.

C3b. Boundedness
$\langle \emptyset, \Omega \rangle \sim \langle \emptyset, \Omega_A \rangle \precsim \langle A, \Omega_A \rangle \precsim \langle \Omega_A, \Omega_A \rangle \sim \langle \Omega, \Omega \rangle$.

$\mathbf{E}_{\precsim^\dagger}$. $\precsim^\dagger$ Extension
$\langle A, \Omega_A \rangle \precsim \langle A', \Omega_A \rangle$ *if and only if* $A \precsim^\dagger A'$.

$\mathbf{E}_{\prec^\dagger}$. $\prec^\dagger$ Extension
If either both $A \prec^\dagger_\Omega B$ and $\Omega_B \precsim^\dagger_\Omega \Omega_A$ or both $\emptyset \prec^\dagger_\Omega A \precsim^\dagger_\Omega B$ and $\Omega_B \prec^\dagger_\Omega \Omega_A$, then $\langle A, \Omega_A \rangle \prec \langle B, \Omega_B \rangle$.

FC. Finite Comparisons
If Ω_A and Ω_B are finite, then $\langle A, \Omega_A \rangle \precsim \langle B, \Omega_B \rangle$ if and only if $\mu_{\Omega_A}(A) \leq \mu_{\Omega_B}(B)$ where $\mu_{\Omega_A} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ and $\mu_{\Omega_B} : \mathcal{P}(\Omega_A) \to \mathbb{Q} \cap [0,1]$ are the unique representations of $\precsim^\dagger_{\Omega_A}$ and $\precsim^\dagger_{\Omega_B}$ in $\langle \mathbb{Q} \cap [0,1], + \rangle$.

ID. Infinitesimal Degrees
If A, B are nonempty and finite sets, Ω_A is infinite, and Ω_B is finite, then $\langle A, \Omega_A \rangle \prec \langle B, \Omega_B \rangle$.

Proof.
C0: $\langle \emptyset, \Omega \rangle \precsim \langle \Omega, \Omega \rangle$ follows from (P). Since Ω is infinite, it suffices to verify that $\langle \Omega, \Omega \rangle \precsim \langle \emptyset, \Omega \rangle$ does not follow from any one of (P), (C), and (I) in isolation. Since $\Omega \not\precsim^\dagger \emptyset$, $\langle \Omega, \Omega \rangle \precsim \langle \emptyset, \Omega \rangle$ does not follow from (P). Since $\Omega \neq \emptyset$, neither does this follows from (C) or (I). Since equivalences between pairs with infinite second parameters are only induced by these, it must be that $\langle \Omega, \Omega \rangle \not\precsim \langle \emptyset, \Omega \rangle$.
C1a: Follows from (P).

C2: By definition.
C3b: Follows from (P), (C), (I), and transitivity.
$E_{\precsim^\dagger}$:
($\Rightarrow$)
Suppose that $\langle A, \Omega_A\rangle \precsim \langle A', \Omega_A\rangle$. If Ω_A is finite, then $A \precsim^\dagger_{\Omega_A} A'$ follows from unique representation in $\langle \mathbb{Q} \cap [0,1], +\rangle$. Suppose further, then, that Ω_A is infinite.

An aside on the the exact structure of $\precsim^{PUCI}$ is useful here. Note first that closing all instances of (U) under transitivity reduces all pairs $\langle B, \Omega_B\rangle$ into equivalence classes where any pair with an infinite Ω_B is alone. The equivalence classes induced by the transitive closure of all (U)-comparisons are further reduced by those $\sim$ relations required by (P), viz. if $B \sim^\dagger_\Omega C$ and $\Omega_C \sim^\dagger_\Omega \Omega_B$, then $\langle B, \Omega_B\rangle \sim \langle C, \Omega_C\rangle$. The antecedent here is equivalent to $B - C$, $C - B$, $\Omega_B - \Omega_C$, and $\Omega_C - \Omega_B$ are all finite with $|B - C| = |C - B|$ and $|\Omega_B - \Omega_C| = |\Omega_C - \Omega_B|$. The reduced equivalence classes are thus the finite n-uniform expansion classes plus finite replacements in each component. If Ω_B is infinite, equivalent pairs at this stage are exclusively induced by (P). Finally, these classes are further reduced by (C) and (I) so that all $\langle \Omega_B, \Omega_B\rangle$ and all $\langle \emptyset, \Omega_B\rangle$ appear in the same class. Finally, pairs equivalent to $\langle B, \Omega_B\rangle$ are either exclusively induced by (P) and (U) ($B \neg \emptyset, \Omega_B$), all induced by (C) ($B = \Omega_B$), or all induced by (I) ($B = \emptyset$) since in the latter cases all equivalences required by (P) and (U) are also required by (C) and (I) respectively.

The only remaining comparisons and the only non-$\sim$ comparisons between equivalence classes are induced by strictly one-sided instances of (P). We show that these are already transitive; whenever $\langle B, \Omega_B\rangle \precsim \langle C, \Omega_C\rangle$ and $\langle C', \Omega_{C'}\rangle \precsim \langle D, \Omega_D\rangle$ by (P) with $\langle C, \Omega_C\rangle$ and $\langle C', \Omega_{C'}\rangle$ in the same equivalence class, there exists $\langle B^*, \Omega_{B^*}\rangle$ and $\langle D^*, \Omega_{D*}\rangle$ equivalent to $\langle B, \Omega_B\rangle$ and $\langle D, \Omega_D\rangle$ respectively such that $\langle B^*, \Omega_{B^*}\rangle \precsim \langle D^*, \Omega_{D^*}\rangle$ by (P). Suppose that $\langle B, \Omega_B\rangle \precsim \langle C, \Omega_C\rangle$ and $\langle C', \Omega_{C'}\rangle \precsim \langle D, \Omega_D\rangle$ by (P) with $\langle C, \Omega_C\rangle$ and $\langle C', \Omega_{C'}\rangle$ equivalent. If $C = \emptyset$, then $B = \emptyset$ as well and thus $\langle \emptyset, \Omega_D\rangle$ is equivalent to $\langle B, \Omega_B\rangle$ with $\langle \emptyset, \Omega_D\rangle \precsim \langle D, \Omega_D\rangle$ by (P). If $C = \Omega_C$, then $C' = \Omega_{C'}$ as well and thus $\langle \Omega_B, \Omega_B\rangle$ is equivalent to $\langle D, \Omega_D\rangle$ with $\langle B, \Omega_B\rangle \precsim \langle \Omega_B, \Omega_B\rangle$ by (P). Suppose then that $C \neq \emptyset, \Omega_C$.

- Ω_B, Ω_C, and Ω_D finite. It follows immediately that $\Omega_{C'}$ is finite as well. Further by $\langle C, \Omega_C\rangle$ and $\langle C', \Omega_{C'}\rangle$ in the same equivalence class, $\frac{|C|}{|\Omega_C|} = \frac{|C'|}{|\Omega_{C'}|}$. Thus,

$$\frac{|B|}{|\Omega_B|} \leq \frac{|C|}{|\Omega_C|} = \frac{|C'|}{|\Omega_{C'}|} \leq \frac{|D|}{|\Omega_D|}.$$

  Taking a $|\Omega_D|$-uniform expansion of Ω_B and a $|\Omega_B|$-uniform expansion of Ω_D, gives a pair $\langle B^*, \Omega_{B^*}\rangle$ and $\langle D^*, \Omega_{D^*}\rangle$ equivalent to $\langle B, \Omega_B\rangle$ and $\langle D, \Omega_D\rangle$ respectively. Further, $|\Omega_{B^*}| = |\Omega_{D^*}|$ and $|B^*| \leq |D^*|$. $\langle B^*, \Omega_{B^*}\rangle \precsim \langle D^*, \Omega_{D^*}\rangle$ follows by (P).

- Ω_B infinite; Ω_C, and Ω_D finite. It follows immediately that $\Omega_{C'}$ is finite as well. Since $|\Omega_B|$ is larger than $|\Omega_{D^*}|$ for any pair $\langle D^*, \Omega_{D^*}\rangle$ equivalent to $\langle D, \Omega_D\rangle$, we need only find an n-uniform expansion of Ω_D with $|D^*| \geq |B|$. Since B is finite (Ω_C is finite and $|B| \leq |C|$ by hypothesis), this is always possible (if D is empty, so too is B; otherwise, take a $|B|$-uniform expansion). $\langle B, \Omega_B\rangle \precsim \langle D^*, \Omega_{D^*}\rangle$ then follows by (P).

- Ω_B finite; Ω_C infinite. Impossible; $\Omega_C \not\precsim^\dagger \Omega_B$.

- Ω_C finite; Ω_D infinite. Impossible; $\Omega_{C'}$ must be finite as well but then $\Omega_D \not\precsim^\dagger \Omega_{C'}$.

- Ω_B and Ω_C infinite; Ω_D finite. Since $\langle C', \Omega_{C'} \rangle \precsim \langle D, \Omega_D \rangle$ by (P), $|C'| \leq |D|$ and thus finite. It follows immediately that C is finite as well. Since $\langle B, \Omega_B \rangle \precsim \langle C, \Omega_C \rangle$ by (P), $|B| \leq |C|$ and thus finite. Since $|\Omega_B|$ is larger than $|\Omega_{D^*}|$ for any pair $\langle D^*, \Omega_{D^*} \rangle$ equivalent to $\langle D, \Omega_D \rangle$, we need only find an n-uniform expansion of Ω_D with $|D^*| \geq |B|$. Since B is finite, this is always possible (if D is empty, so too is B; otherwise, take a $|B|$-uniform expansion). $\langle B, \Omega_B \rangle \precsim \langle D^*, \Omega_{D^*} \rangle$ then follows by (P).

- Ω_B, Ω_C, and Ω_D infinite. $\langle C, \Omega_C \rangle$ and $\langle C', \Omega_{C'} \rangle$ must be equivalent by (P). $B \precsim^\dagger D$ and $\Omega_D \precsim^\dagger \Omega_B$ follows by the transitivity of $\precsim^\dagger$. $\langle B, \Omega_B \rangle \precsim \langle D, \Omega_D \rangle$ then follows by (P).

Returning to $\langle A, \Omega_A \rangle \precsim^{PUCI} \langle A', \Omega_A \rangle$ with Ω_A infinite, it must thus be either that $\langle A, \Omega_A \rangle \precsim^{PUCI} \langle A', \Omega_A \rangle$ is a result of closing under those $\sim$ relations required by (P), those required by (C), those required by (I), or by a strictly one-sided instance of (P). In all four cases, $A \precsim^\dagger_\Omega A'$ and thus $A \precsim^\dagger_{\Omega_A} A'$ follows immediately.

($\Leftarrow$)
Follows immediately from $\precsim^{PUCI}$ an extension of $\precsim^{PU}$.

$E_{\prec^\dagger}$: Suppose that either $A \prec^\dagger_\Omega B$ and $\Omega_B \precsim^\dagger_\Omega \Omega_A$ or $A \precsim^\dagger_\Omega B$ and $\Omega_B \prec^\dagger_\Omega \Omega_A$. We proceed by cases.

- Case 1: $A \prec^\dagger_\Omega B$ and $\Omega_B \precsim^\dagger_\Omega \Omega_A$. $\langle A, \Omega_A \rangle \precsim \langle B, \Omega_B \rangle$ follows immediately by (P). It remains only to show that $\langle B, \Omega_B \rangle \not\precsim \langle A, \Omega_A \rangle$. $\langle B, \Omega_B \rangle$ and $\langle A, \Omega_A \rangle$ are not equivalent by (P) alone since $A \prec^\dagger B$. Neither could they be equivalent by (I) since this requires $A = B = \emptyset$. (C) likewise fails since it requires both $A = \Omega_A$ and $B = \Omega_B$ from which it follows that $B \precsim^\dagger A$. If Ω_A or Ω_B are infinite, we thus have that $\langle B, \Omega_B \rangle \not\precsim \langle A, \Omega_A \rangle$ (recall the discussion of $\precsim^{PUCI}$ above). Supposing that both Ω_A and Ω_B are finite, then $\langle B, \Omega_B \rangle \not\precsim \langle A, \Omega_A \rangle$ by (FC) (proved below) and the observation that $|A| < |B|$ while $|\Omega_B| \leq |\Omega_A|$.

- Case 2: $A \precsim^\dagger_\Omega B$ and $\Omega_B \prec^\dagger_\Omega \Omega_A$. $\langle A, \Omega_A \rangle \precsim \langle B, \Omega_B \rangle$ follows immediately by (P). It remains only to show that $\langle B, \Omega_B \rangle \not\precsim \langle A, \Omega_A \rangle$. $\langle B, \Omega_B \rangle$ and $\langle A, \Omega_A \rangle$ are not equivalent by (P) alone since $\Omega_B \prec^\dagger \Omega_A$. Neither could they be equivalent by (I) since this requires $A = B = \emptyset$. (C) likewise fails since it requires both $A = \Omega_A$ and $B = \Omega_B$ from which it follows that $\Omega_A \precsim^\dagger \Omega_B$. If Ω_A or Ω_B are infinite, we thus have that $\langle B, \Omega_B \rangle \not\precsim \langle A, \Omega_A \rangle$ (recall the discussion of $\precsim^{PUCI}$ above). Supposing that both Ω_A and Ω_B are finite, then $\langle B, \Omega_B \rangle \not\precsim \langle A, \Omega_A \rangle$ by (FC) (proved below) and the observation that $0 < |A| \leq |B|$ while $|\Omega_B| < |\Omega_A|$.

(FC): Follows immediately from the corresponding result for $\precsim^{PU}$ by noting that $\precsim^{PU}$ and $\precsim^{PUCI}$ are equivalent over finite sets (all instances of (C) and (I) here follow from (U), (P), and transitivity).

(ID): Suppose A, B are nonempty and finite sets, Ω_A is infinite, and Ω_B is finite. Since Ω_B is finite, construct a $|A|$-uniform extension of Ω_B to $\Omega_{B'}$. Let B' be the subset of $\Omega_{B'}$ which corresponds to B. By (U), $\langle B, \Omega_B \rangle \sim \langle B', \Omega'_B \rangle$. Since $B \neq \emptyset$, $\emptyset \prec^{\dagger} A \precsim^{\dagger} B'$ and $\Omega_{B'} \prec^{\dagger} \Omega_A$. By $E_{\prec^{\dagger}}$, $\langle A, \Omega_A \rangle \prec \langle B', \Omega_{B'} \rangle$ and thus $\langle A, \Omega_A \rangle \prec \langle B, \Omega_B \rangle$ as desired. $\square$